GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 12, 2003, 11:12:01 ; Search time 40 Seconds
(without alignments)
16.830 Million cell updates/sec

Title: US-09-939-293A-19_COPY_56_62
Perfect score: 33
Sequence: 1 AVPIAQK 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	30	90.9	602	2	AB0024	probable potassium
2	29	87.9	211	2	H72652	hypothetical prote
3	29	87.9	402	2	D70602	probable arginine
4	29	87.9	420	2	S77102	hypothetical prote
5	29	87.9	437	2	C75085	hypothetical prote
6	29	87.9	718	2	AE1832	ATP-dependent DNA
7	28	84.8	145	2	G84120	ribose 5-phosphate
8	28	84.8	301	2	G71206	tryptophan-tRNA li
9	28	84.8	323	2	AG2128	hypothetical prote
10	28	84.8	385	2	C75020	tryptophanyl-tRNA
11	28	84.8	390	1	W2WLRB	E2 protein - cotto
12	28	84.8	445	2	AD2184	hypothetical prote
13	28	84.8	476	2	D87503	pyruvate kinase [i
14	28	84.8	656	2	B82056	glutathione-regula
15	28	84.8	845	2	T40955	hypothetical prote
16	28	84.8	867	2	S72842	methionine synthas
17	28	84.8	1068	2	F84614	probable kinesin h
18	28	84.8	1206	2	E87072	hypothetical prote
19	28	84.8	1581	1	VGWJBV	peplomer glycoprot
20	27	81.8	92	2	E47754	yceA protein homol
21	27	81.8	115	2	G81438	probable periplasm
22	27	81.8	157	2	B81090	conserved hypothet
23	27	81.8	157	2	D81850	hypothetical prote
24	27	81.8	185	2	A43309	outer membrane pro
25	27	81.8	228	2	E86849	glutamate ABC tran
26	27	81.8	232	2	T38619	hypothetical prote
27	27	81.8	256	2	AE2068	hypothetical prote
28	27	81.8	268	2	F64024	hypothetical prote
29	27	81.8	289	2	S77303	hypothetical prote
30	27	81.8	307	2	T15037	DNA-binding protei
31	27	81.8	307	2	T15038	DNA-binding protei
32	27	81.8	319	2	A86777	conserved hypothet
33	27	81.8	336	2	AC1979	ferric iron-bindin
34	27	81.8	375	2	T03593	leucoanthocyanidin
35	27	81.8	407	2	I55525	myeloid cell nucle
36	27	81.8	407	2	G84309	hypothetical prote
37	27	81.8	443	2	D71058	hypothetical prote
38	27	81.8	443	2	AD1831	hypothetical prote
39	27	81.8	447	2	S53309	n-6 fatty acid des
40	27	81.8	447	2	T47729	hypothetical prote
41	27	81.8	503	2	T35053	probable solute-bi
42	27	81.8	508	2	S73430	glycerol kinase gl
43	27	81.8	541	2	T19304	hypothetical prote
44	27	81.8	577	2	S56445	hypothetical 64.8K
45	27	81.8	577	2	F91278	hypothetical prote

ALIGNMENTS

RESULT 1
AB0024
probable potassium-efflux system protein [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AB0024
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AB0024
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-602 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89052.1; PID:g15978292; GSPDB:GN00175
C;Genetics:
A;Gene: kefB
C;Superfamily: glutathione-regulated potassium efflux system protein kefC

Query Match 90.9%; Score 30; DB 2; Length 602;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
||||||:
Db 20 AVPIAQR 26

RESULT 2
H72652
hypothetical protein APE0653 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 08-Sep-2000
C;Accession: H72652
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: H72652
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-211 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79624.1; PID:d1043410; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0653
C;Superfamily: Aeropyrum pernix hypothetical protein APE0653

Query Match 87.9%; Score 29; DB 2; Length 211;

```
  Best Local Similarity   71.4%;  Pred. No. 25;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:||:
Db         31 AVPVAQR 37


RESULT 3
D70602
probable arginine deiminase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70602
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: D70602
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-402 <COL>
A;Cross-references: GB:Z94752; GB:AL123456; NID:g3261731; PIDN:CAB08144.1; PID:g2052136
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: arcA
C;Superfamily: arginine deiminase arcA

  Query Match             87.9%;  Score 29;  DB 2;  Length 402;
  Best Local Similarity   85.7%;  Pred. No. 50;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||||:
Db        257 AVPIAQQ 263


RESULT 4
S77102
hypothetical protein slr1865 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S77102
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S77102
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-420 <KAN>
A;Cross-references: EMBL:D90908; GB:AB001339; NID:g1652725; PIDN:BAA17660.1; PID:d101839
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

  Query Match             87.9%;  Score 29;  DB 2;  Length 420;
  Best Local Similarity   85.7%;  Pred. No. 53;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||||:
Db        115 AVPIAQQ 121


RESULT 5
C75085
hypothetical protein PAB1660 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: C75085
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: C75085
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <KAW>
A;Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB49984.1; PID:g545
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1660
C;Superfamily: conserved hypothetical protein HI0125

  Query Match             87.9%;  Score 29;  DB 2;  Length 437;
  Best Local Similarity   71.4%;  Pred. No. 55;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:||:||
Db        128 AIPISQK 134


RESULT 6
AE1832
ATP-dependent DNA helicase [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AE1832
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE1832
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-718 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB77729.1; PID:g17135183; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr0205
C;Superfamily: recQ protein; recQ helicase homology

  Query Match             87.9%;  Score 29;  DB 2;  Length 718;
  Best Local Similarity   71.4%;  Pred. No. 95;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              :||:|||
Db        509 SVPVAQK 515


RESULT 7
G84120
ribose 5-phosphate epimerase (pentose phosphate) BH3767 [imported] - Bacillus halodur
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: G84120
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: G84120
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-145 <STO>
A;Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BAB07486.1; GSPDB:G
```

```
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH3767
C;Superfamily: galactoside O-acetyltransferase

  Query Match             84.8%;  Score 28;  DB 2;  Length 145;
  Best Local Similarity   57.1%;  Pred. No. 29;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:|:|:|
Db         46 AIPVAEK 52


RESULT 8
G71206
tryptophan-tRNA ligase (EC 6.1.1.2) - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 03-Jun-2002
C;Accession: G71206
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: G71206
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-301 <KAW>
A;Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31046.1; PID:g3258363
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1921
C;Superfamily: yeast tyrosine-tRNA ligase
C;Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

  Query Match             84.8%;  Score 28;  DB 2;  Length 301;
  Best Local Similarity   71.4%;  Pred. No. 64;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:|||:|
Db         80 AIPIAKK 86


RESULT 9
AG2128
hypothetical protein alr2582 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AG2128
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2128
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-323 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB74281.1; PID:g17131674; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr2582

  Query Match             84.8%;  Score 28;  DB 2;  Length 323;
  Best Local Similarity   85.7%;  Pred. No. 69;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||| ||
Db        108 AVPIIQK 114


RESULT 10
C75020
tryptophanyl-tRNA synthetase (trps) PAB1111 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: C75020
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: C75020
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-385 <KAW>
A;Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50601.1; PID:g545
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: trpS; PAB1111
C;Superfamily: mammalian tryptophan-tRNA ligase; amino acid-tRNA ligase repeat homolo

  Query Match             84.8%;  Score 28;  DB 2;  Length 385;
  Best Local Similarity   71.4%;  Pred. No. 84;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:|||:|
Db        165 AIPIAKK 171


RESULT 11
W2WLRB
E2 protein - cottontail rabbit papillomavirus
C;Species: cottontail rabbit papillomavirus
C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 24-Feb-1994
C;Accession: A03671
R;Giri, I.; Danos, O.; Yaniv, M.
Proc. Natl. Acad. Sci. U.S.A. 82, 1580-1584, 1985
A;Title: Genomic structure of the cottontail rabbit (Shope) papillomavirus.
A;Reference number: A94027; MUID:85166175; PMID:2984661
A;Accession: A03671
A;Molecule type: DNA
A;Residues: 1-390 <GIR>
C;Superfamily: papillomavirus E2 protein
C;Keywords: early protein

  Query Match             84.8%;  Score 28;  DB 1;  Length 390;
  Best Local Similarity   85.7%;  Pred. No. 85;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||| |||
Db        222 AVPAAQK 228


RESULT 12
AD2184
hypothetical protein alr3027 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AD2184
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2184
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-445 <KUR>
```

```
A;Cross-references: GB:BA000019; PIDN:BAB74726.1; PID:g17132121; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3027
C;Superfamily: conserved hypothetical protein HI1029

  Query Match             84.8%;  Score 28;  DB 2;  Length 445;
  Best Local Similarity   83.3%;  Pred. No. 98;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              ||:|||
Db        362 VPVAQK 367


RESULT 13
D87503
pyruvate kinase [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: D87503
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: D87503
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-476 <STO>
A;Cross-references: GB:AE005673; NID:g13423528; PIDN:AAK24024.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2051
C;Superfamily: pyruvate kinase

  Query Match             84.8%;  Score 28;  DB 2;  Length 476;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              ||:|||
Db        253 VPVAQK 258


RESULT 14
B82056
glutathione-regulated potassium-efflux system protein KefB VC2606 [imported] - Vibrio ch
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: B82056
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: B82056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-656 <HEI>
A;Cross-references: GB:AE004327; GB:AE003852; NID:g9657185; PIDN:AAF95747.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC2606
A;Map position: 1
C;Superfamily: glutathione-regulated potassium efflux system protein kefC

  Query Match             84.8%;  Score 28;  DB 2;  Length 656;
  Best Local Similarity   71.4%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
                |||:||:
Db         78 AVPLAQR 84


RESULT 15
T40955
hypothetical protein SPCC1393.07c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C;Accession: T40955
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, February 1999
A;Reference number: Z21940
A;Accession: T40955
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-845 <WOO>
A;Cross-references: EMBL:AL035592; PIDN:CAB38163.1; GSPDB:GN00068; SPDB:SPCC1393.07c
A;Experimental source: strain 972h-; cosmid c1393
C;Genetics:
A;Gene: SPDB:SPCC1393.07c
A;Map position: 3
A;Introns: 408/1
C;Superfamily: Schizosaccharomyces pombe hypothetical protein SPCC1393.07c

  Query Match             84.8%;  Score 28;  DB 2;  Length 845;
  Best Local Similarity   83.3%;  Pred. No. 2e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              :|||||
Db         46 IPIAQK 51


Search completed: September 12, 2003, 11:17:01
Job time : 43 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 12, 2003, 11:09:46 ; Search time 95 Seconds
(without alignments)
19.014 Million cell updates/sec

Title: US-09-939-293A-19_COPY_56_62
Perfect score: 33
Sequence: 1 AVPIAQK 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	100.0	157	11	Q8R1D8	Q8r1d8 mus musculu
2	33	100.0	344	11	Q8BW93	Q8bw93 mus musculu
3	31	93.9	1010	5	Q9VFG4	Q9vfg4 drosophila
4	30	90.9	359	4	Q8TBM0	Q8tbm0 homo sapien
5	30	90.9	578	4	Q9H8K8	Q9h8k8 homo sapien
6	30	90.9	578	4	Q96FU5	Q96fu5 homo sapien
7	30	90.9	602	16	Q8ZJC4	Q8zjc4 yersinia pe
8	30	90.9	635	12	Q8JZ11	Q8jz11 beet wester
9	29	87.9	44	5	Q9NFP7	Q9nfp7 ceratitis r
10	29	87.9	197	1	Q977W0	Q977w0 uncultured
11	29	87.9	198	1	Q977V8	Q977v8 uncultured
12	29	87.9	198	1	Q977V9	Q977v9 uncultured
13	29	87.9	211	17	Q9YEC2	Q9yec2 aeropyrum p
14	29	87.9	366	1	Q977M5	Q977m5 uncultured
15	29	87.9	371	11	Q8BW45	Q8bw45 mus musculu
16	29	87.9	408	16	Q8XHY1	Q8xhy1 clostridium
17	29	87.9	420	16	P73615	P73615 synechocyst
18	29	87.9	434	2	Q9F9X2	Q9f9x2 streptococc
19	29	87.9	434	2	Q9ETV8	Q9etv8 streptococc
20	29	87.9	434	2	Q9EU98	Q9eu98 streptococc
21	29	87.9	434	16	Q8E7W4	Q8e7w4 streptococc
22	29	87.9	434	16	Q8E2F6	Q8e2f6 streptococc
23	29	87.9	437	1	Q56470	Q56470 uncultured
24	29	87.9	437	17	Q9UZS2	Q9uzs2 pyrococcus
25	29	87.9	718	16	Q8Z093	Q8z093 anabaena sp
26	29	87.9	1323	10	Q8LK00	Q8lk00 sorghum bic
27	29	87.9	1340	16	Q8PEP6	Q8pep6 xanthomonas
28	29	87.9	3075	13	Q8AW10	Q8aw10 brachydanio
29	29	87.9	4641	4	O75592	O75592 homo sapien
30	28	84.8	44	5	Q95ZI1	Q95zi1 ceratitis r
31	28	84.8	145	16	Q9K6G2	Q9k6g2 bacillus ha
32	28	84.8	147	12	Q85062	Q85062 cottontail
33	28	84.8	285	16	Q8DWG7	Q8dwg7 streptococc
34	28	84.8	300	5	Q9NH08	Q9nh08 agrotis ips
35	28	84.8	315	11	Q8CIF0	Q8cif0 mus musculu
36	28	84.8	323	16	Q8YTY1	Q8yty1 anabaena sp
37	28	84.8	324	16	Q8ETN2	Q8etn2 oceanobacil
38	28	84.8	362	11	Q8BR14	Q8br14 mus musculu
39	28	84.8	390	12	Q9IES1	Q9ies1 cottontail
40	28	84.8	390	12	Q9ICK8	Q9ick8 cottontail
41	28	84.8	414	16	Q98DE7	Q98de7 rhizobium l
42	28	84.8	420	17	Q979W5	Q979w5 thermoplasm
43	28	84.8	445	16	Q8YSQ7	Q8ysq7 anabaena sp
44	28	84.8	476	16	Q9A6N6	Q9a6n6 caulobacter
45	28	84.8	482	5	Q9Y066	Q9y066 trypanosoma

ALIGNMENTS

RESULT 1
Q8R1D8
ID Q8R1D8 PRELIMINARY; PRT; 157 AA.
AC Q8R1D8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to RIKEN cDNA 0610041G12 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC024780; AAH24780.1; -.
SQ SEQUENCE 157 AA; 17799 MW; 0F67319F05EAC6E7 CRC64;

Query Match 100.0%; Score 33; DB 11; Length 157;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 54 AVPIAQK 60

RESULT 2
Q8BW93
ID Q8BW93 PRELIMINARY; PRT; 344 AA.
AC Q8BW93;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Weakly similar to G protein-coupled receptor C5L2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Lung;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK053187; BAC35303.1; -.
SQ   SEQUENCE   344 AA;  38198 MW;  508FFD23F01B31C8 CRC64;

  Query Match           100.0%;  Score 33;  DB 11;  Length 344;
  Best Local Similarity  100.0%;  Pred. No. 22;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db         97 AVPIAQK 103


RESULT 3
Q9VFG4
ID   Q9VFG4      PRELIMINARY;      PRT;  1010 AA.
AC   Q9VFG4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG7362 protein.
GN   CG7362.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003706; AAF55096.2; -.
DR   HSSP; P14178; 1E0T.
DR   FlyBase; FBgn0038258; CG7362.
DR   InterPro; IPR001697; Pyruvate_kinase.
DR   Pfam; PF00224; PK; 1.
DR   Pfam; PF02887; PK_C; 1.
DR   ProDom; PD001009; Pyruvate_kinase; 1.
DR   TIGRFAMs; TIGR01064; pyruv_kin; 1.
DR   PROSITE; PS00110; PYRUVATE_KINASE; 1.
SQ   SEQUENCE   1010 AA;  113769 MW;  9D0415C4C651C15A CRC64;

  Query Match           93.9%;  Score 31;  DB 5;  Length 1010;
  Best Local Similarity  85.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:|||
Db        357 AVPLAQK 363


RESULT 4
Q8TBM0
ID   Q8TBM0      PRELIMINARY;      PRT;  359 AA.
AC   Q8TBM0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RA   Strausberg R.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC022341; AAH22341.1; -.
DR   InterPro; IPR001278; Arg_tRNA-synt_1c.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00750; tRNA-synt_1d; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   359 AA;  40568 MW;  ED0615B69F3C0617 CRC64;

  Query Match             90.9%;  Score 30;  DB 4;  Length 359;
  Best Local Similarity   85.7%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||:||
Db         30 AVPISQK 36


RESULT 5
Q9H8K8
ID   Q9H8K8      PRELIMINARY;      PRT;   578 AA.
AC   Q9H8K8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ13488.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA   Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA   Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA   Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK023550; BAB14608.1; -.
DR   HSSP; Q05506; 1BS2.
DR   InterPro; IPR001278; Arg_tRNA-synt_1c.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00750; tRNA-synt_1d; 1.
DR   PRINTS; PR01038; TRNASYNTHARG.
DR   TIGRFAMs; TIGR00456; argS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   578 AA;  65487 MW;  68BE7107DA923C4B CRC64;

  Query Match             90.9%;  Score 30;  DB 4;  Length 578;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||:||
Db         30 AVPISQK 36


RESULT 6
Q96FU5
ID   Q96FU5      PRELIMINARY;      PRT;   578 AA.
AC   Q96FU5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC010420; AAH10420.1; -.
DR   InterPro; IPR001278; Arg_tRNA-synt_1c.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00750; tRNA-synt_1d; 1.
DR   PRINTS; PR01038; TRNASYNTHARG.
DR   TIGRFAMs; TIGR00456; argS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   578 AA;  65533 MW;  17F28D28E8805284 CRC64;

  Query Match             90.9%;  Score 30;  DB 4;  Length 578;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||:||
Db         30 AVPISQK 36


RESULT 7
Q8ZJC4
ID   Q8ZJC4      PRELIMINARY;      PRT;   602 AA.
AC   Q8ZJC4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Probable potassium-efflux system protein (K+ efflux, NEM-activable
DE   K+/H+ antiporter).
GN   KEFB OR YPO0191 OR Y3972.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
DR   EMBL; AJ414141; CAC89052.1; -.
DR   EMBL; AE014001; AAM87516.1; -.
DR   InterPro; IPR004771; K_eff.
DR   InterPro; IPR006153; Na_H_porter.
DR   InterPro; IPR006036; TrkA_Kuptake.
DR   InterPro; IPR003148; TrkA_N.
DR   Pfam; PF00999; Na_H_Exchanger; 1.
```

```
DR   Pfam; PF02254; TrkA-N; 1.
DR   PRINTS; PR00335; KUPTAKETRKA.
DR   TIGRFAMs; TIGR00932; 2a37; 1.
KW   Complete proteome.
SQ   SEQUENCE   602 AA;  66328 MW;  3166D7C15AE7C80A CRC64;

  Query Match             90.9%;  Score 30;  DB 16;  Length 602;
  Best Local Similarity   85.7%;  Pred. No. 2e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||||:
Db         20 AVPIAQR 26


RESULT 8
Q8JZ11
ID   Q8JZ11      PRELIMINARY;      PRT;   635 AA.
AC   Q8JZ11;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   RNA-dependent RNA polymerase P1 protein.
OS   Beet western yellows virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC   Polerovirus.
OX   NCBI_TaxID=12042;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USA;
RA   Beuve M., Lemaire O.;
RT   "Sugar beet-infecting beet western yellows virus (BWYV)-USA strain
RT   represents a distinct Polerovirus species.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF473561; AAM22680.1; -.
DR   InterPro; IPR000382; Luteo_ORF2.
DR   Pfam; PF02122; Luteo_ORF2; 1.
DR   PRINTS; PR00913; LVIRUSORF2.
KW   RNA-directed RNA polymerase.
SQ   SEQUENCE   635 AA;  69341 MW;  9894FA7D2ADED9B0 CRC64;

  Query Match             90.9%;  Score 30;  DB 12;  Length 635;
  Best Local Similarity   85.7%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||:||
Db        476 AVPISQK 482


RESULT 9
Q9NFP7
ID   Q9NFP7      PRELIMINARY;      PRT;    44 AA.
AC   Q9NFP7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Ceratotoxin 1 precursor (Fragment).
GN   CR1.
OS   Ceratitis rosa (Natal fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Tephritoidea; Tephritidae; Ceratitis.
OX   NCBI_TaxID=56958;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rosetto M.;
RT   "Evolution of the ceratotoxin gene family in the medfly Ceratitis
RT   capitata and the Natal fruit fly Ceratitis rosa.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ272450; CAB75957.1; -.
KW   Signal.
FT   NON_TER       1      1
FT   SIGNAL        <1     17       POTENTIAL.
FT   CHAIN         30    >44       CERATOTOXIN 1.
FT   NON_TER       44     44
SQ   SEQUENCE   44 AA;  4676 MW;  C81B7D0C0D4AB270 CRC64;

  Query Match             87.9%;  Score 29;  DB 5;  Length 44;
  Best Local Similarity   85.7%;  Pred. No. 23;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||:|
Db         38 AVPIAKK 44


RESULT 10
Q977W0
ID   Q977W0      PRELIMINARY;      PRT;   197 AA.
AC   Q977W0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Glutamate semialdehyde aminotransferase (Fragment).
OS   uncultured crenarchaeote 15G10.
OC   Archaea; Crenarchaeota; environmental samples;
OC   marine archaeal group 1.
OX   NCBI_TaxID=166582;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=15G10;
RA   Beja O.;
RT   "Comparative genomic analysis of coexisting archaeal genetic variants
RT   in an antarctic marine microbial assemblage.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF393304; AAK76997.1; -.
DR   InterPro; IPR005814; Aminotrans_3.
DR   Pfam; PF00202; aminotran_3; 2.
KW   Aminotransferase; Transferase.
FT   NON_TER     197    197
SQ   SEQUENCE   197 AA;  22017 MW;  5E6A833C898E9DF2 CRC64;

  Query Match             87.9%;  Score 29;  DB 1;  Length 197;
  Best Local Similarity   71.4%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:|:|
Db         46 AVPVAEK 52


RESULT 11
Q977V8
ID   Q977V8      PRELIMINARY;      PRT;   198 AA.
AC   Q977V8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Glutamate semialdehyde aminotransferase (Fragment).
OS   uncultured crenarchaeote 83A10.
OC   Archaea; Crenarchaeota; environmental samples;
OC   marine archaeal group 1.
OX   NCBI_TaxID=166585;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=83A10;
RA   Beja O.;
RT   "Comparative genomic analysis of coexisting archaeal genetic variants
RT   in an antarctic marine microbial assemblage.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF393307; AAK76999.1; -.
DR   InterPro; IPR005814; Aminotrans_3.
DR   Pfam; PF00202; aminotran_3; 2.
KW   Aminotransferase; Transferase.
FT   NON_TER     198    198
```

```
SQ   SEQUENCE   198 AA;  22087 MW;  00BA012F6B8F16D8 CRC64;

  Query Match             87.9%;  Score 29;  DB 1;  Length 198;
  Best Local Similarity   71.4%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:|:|
Db         46 AVPVAEK 52


RESULT 12
Q977V9
ID   Q977V9      PRELIMINARY;      PRT;   198 AA.
AC   Q977V9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Glutamate semialdehyde aminotransferase (Fragment).
OS   uncultured crenarchaeote 31B02.
OC   Archaea; Crenarchaeota; environmental samples;
OC   marine archaeal group 1.
OX   NCBI_TaxID=166584;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=31B02;
RA   Beja O.;
RT   "Comparative genomic analysis of coexisting archaeal genetic variants
RT   in an antarctic marine microbial assemblage.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF393306; AAK76998.1; -.
DR   InterPro; IPR005814; Aminotrans_3.
DR   Pfam; PF00202; aminotran_3; 2.
KW   Aminotransferase; Transferase.
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  22029 MW;  00B9CDE06843D9D8 CRC64;

  Query Match             87.9%;  Score 29;  DB 1;  Length 198;
  Best Local Similarity   71.4%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:|:|
Db         46 AVPVAEK 52


RESULT 13
Q9YEC2
ID   Q9YEC2      PRELIMINARY;      PRT;   211 AA.
AC   Q9YEC2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein APE0653.
GN   APE0653.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC   Desulfurococcaceae; Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000060; BAA79624.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   211 AA;  23327 MW;  2D603212785A2880 CRC64;

  Query Match             87.9%;  Score 29;  DB 17;  Length 211;
  Best Local Similarity   71.4%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:||:
Db         31 AVPVAQR 37


RESULT 14
Q977M5
ID   Q977M5      PRELIMINARY;      PRT;   366 AA.
AC   Q977M5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Glutamate-1-semialdehyde aminotransferase.
OS   uncultured crenarchaeote 74A4.
OC   Archaea; Crenarchaeota; environmental samples;
OC   marine archaeal group 1.
OX   NCBI_TaxID=166279;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Beja O., Koonin E.V., Aravind L., Taylor L.T., Seiz H., Stein J.L.,
RA   Bensen D.C., Feldman R.A., Swanson R.V., DeLong E.F.;
RT   "Comparative genomic analysis of coexisting archaeal genetic variants
RT   in an Antarctic marine microbial assemblage.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF393466; AAK96083.1; -.
DR   InterPro; IPR005814; Aminotrans_3.
DR   Pfam; PF00202; aminotran_3; 1.
KW   Aminotransferase; Transferase.
SQ   SEQUENCE   366 AA;  40232 MW;  35ADA0D3DDF428E9 CRC64;

  Query Match             87.9%;  Score 29;  DB 1;  Length 366;
  Best Local Similarity   71.4%;  Pred. No. 2e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:|:|
Db         46 AVPVAEK 52


RESULT 15
Q8BW45
ID   Q8BW45      PRELIMINARY;      PRT;   371 AA.
AC   Q8BW45;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to similar to NICE-5 protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Ovary;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK054382; BAC35757.1; -.
SQ   SEQUENCE   371 AA;  42237 MW;  BF0DA7D1B3045C05 CRC64;

  Query Match             87.9%;  Score 29;  DB 11;  Length 371;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 VPIAQK 7
              ||||||
Db        125 VPIAQK 130
```

Search completed: September 12, 2003, 11:15:38
Job time : 98 secs

OM protein - protein search, using sw model

Run on: September 12, 2003, 10:57:21 ; Search time 22 Seconds
(without alignments)
14.963 Million cell updates/sec

Title: US-09-939-293A-19_COPY_56_62
Perfect score: 33
Sequence: 1 AVPIAQK 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	100.0	237	1	SMAC_MOUSE	Q9jiq3 mus musculu
2	33	100.0	239	1	SMAC_HUMAN	Q9nr28 homo sapien
3	30	90.9	736	1	EF2_SULTO	Q975h5 sulfolobus
4	29	87.9	71	1	CERD_CERCA	O17513 ceratitis c
5	29	87.9	402	1	ARCA_MYCTU	O05585 mycobacteri
6	28	84.8	385	1	SYW_PYRAB	Q9uy11 pyrococcus
7	28	84.8	385	1	SYW_PYRFU	Q8u453 pyrococcus
8	28	84.8	386	1	SYW_PYRHO	O59584 pyrococcus
9	28	84.8	390	1	VE2_CRPVK	P03121 cottontail
10	28	84.8	716	1	BAC2_MOUSE	P97303 mus musculu
11	28	84.8	841	1	BAC2_HUMAN	Q9byv9 homo sapien
12	28	84.8	1206	1	METH_MYCLE	Q49775 mycobacteri
13	28	84.8	1581	1	VGLP_BEV	P23052 berne virus
14	28	84.8	2193	1	POLG_CX16G	Q65900 c genome po
15	28	84.8	2193	1	POLG_HE71M	Q66479 h genome po
16	27	81.8	268	1	YC73_HAEIN	P44150 haemophilus
17	27	81.8	319	1	YMDE_LACLA	Q9cg88 lactococcus
18	27	81.8	337	1	RLA0_SULTO	Q971j2 sulfolobus
19	27	81.8	407	1	MNDA_HUMAN	P41218 homo sapien
20	27	81.8	447	1	FD6C_SPIOL	P48629 spinacia ol
21	27	81.8	508	1	GLPK_MYCPN	P75064 mycoplasma
22	27	81.8	577	1	YTFM_ECOLI	P39320 escherichia
23	27	81.8	635	1	DXS_ANASP	Q8yz80 anabaena sp
24	27	81.8	883	1	HSS2_HUMAN	P52849 homo sapien
25	27	81.8	883	1	HSS2_MOUSE	P52850 mus musculu
26	27	81.8	1237	1	POL4_DROME	P10394 drosophila
27	27	81.8	1284	1	NRCA_CHICK	P35331 gallus gall
28	26	78.8	218	1	CLD5_HUMAN	O00501 homo sapien
29	26	78.8	254	1	MOTA_AQUAE	O67122 aquifex aeo
30	26	78.8	266	1	TRPA_LACCA	P17166 lactobacill
31	26	78.8	271	1	ALLR_ECOLI	P77734 escherichia
32	26	78.8	311	1	OSTP_RABIT	P31097 oryctolagus
33	26	78.8	314	1	OSTP_HUMAN	P10451 homo sapien
34	26	78.8	399	1	FTSW_BUCAI	P57312 buchnera ap
35	26	78.8	443	1	YJJJ_ECOLI	P39410 escherichia
36	26	78.8	480	1	GLGA_PASMU	Q9cn91 pasteurella
37	26	78.8	540	1	CH60_STRPN	Q97nv4 streptococc
38	26	78.8	559	1	TRAP_PLAFA	P16893 plasmodium
39	26	78.8	574	1	HMD3_SOLTU	Q41438 solanum tub
40	26	78.8	607	1	PESC_SCHPO	O60164 schizosacch
41	26	78.8	625	1	GCKR_HUMAN	Q14397 homo sapien
42	26	78.8	626	1	GCKR_RAT	Q07071 rattus norv
43	26	78.8	638	1	DNAK_PSESG	Q9wwg9 pseudomonas
44	26	78.8	670	1	CACP_YEAST	P32796 saccharomyc
45	26	78.8	765	1	PHYA_ANASP	Q91cc2 anabaena sp

ALIGNMENTS

RESULT 1
SMAC_MOUSE
ID SMAC_MOUSE STANDARD; PRT; 237 AA.
AC Q9JIQ3; Q9CZD1; Q9DCD3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Smac protein, mitochondrial precursor (Second mitochondria-derived
DE activator of caspase) (Direct IAP binding protein with low pI).
GN SMAC OR DIABLO.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RC STRAIN=BALB/c; TISSUE=Kidney;
RX MEDLINE=20383537; PubMed=10929712;
RA Verhagen A.M., Ekert P.G., Pakusch M., Silke J., Connolly L.M.,
RA Reid G.E., Moritz R.L., Simpson R.J., Vaux D.L.;
RT "Identification of DIABLO, a mammalian protein that promotes apoptosis
RT by binding to and antagonizing IAP proteins.";
RL Cell 102:43-53(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C.,
RA Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H.,
RA Kohtsuki S.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: PROMOTES APOPTOSIS BY ACTIVATING CASPASES IN THE
CC CYTOCHROME C/APAF-1/CASPASE-9 PATHWAY. ACTS BY OPPOSING THE
CC INHIBITORY ACTIVITY OF INHIBITOR OF APOPTOSIS PROTEINS (IAP).
CC -!- SUBUNIT: Homodimer. Interacts with BIRC2, BIRC3, BIRC4/XIAP and
CC BIRC7 (By similarity).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL BUT RELEASED INTO THE CYTOSOL
CC WHEN CELLS UNDERGO APOPTOSIS.
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION FOUND IN HEART, LIVER,

```
CC       KIDNEY AND TESTIS.
CC   -!- DOMAIN: The mature N-terminus mediates interaction with
CC       BIRC4/XIAP (By similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF203914; AAF82190.1; -.
DR   EMBL; AK012760; BAB28450.1; -.
DR   EMBL; AK002887; -; NOT_ANNOTATED_CDS.
DR   HSSP; Q9NR28; 1FEW.
DR   MGD; MGI:1913843; 0610041G12Rik.
KW   Transit peptide; Mitochondrion; Apoptosis.
FT   TRANSIT       1     53       MITOCHONDRION (BY SIMILARITY).
FT   CHAIN        54    237       SMAC PROTEIN.
FT   SITE         54     58       IAP-BINDING MOTIF (BY SIMILARITY).
FT   CONFLICT     64     64       H -> Q (IN REF. 2).
SQ   SEQUENCE   237 AA;  26829 MW;  E53E6F04F1C390A1 CRC64;

  Query Match             100.0%;  Score 33;  DB 1;  Length 237;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db         54 AVPIAQK 60


RESULT 2
SMAC_HUMAN
ID   SMAC_HUMAN     STANDARD;      PRT;   239 AA.
AC   Q9NR28; Q96LV0; Q9BT11; Q9HAV6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Smac protein, mitochondrial precursor (Second mitochondria-derived
DE   activator of caspase) (Direct IAP binding protein with low pI).
GN   SMAC OR DIABLO.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), PARTIAL SEQUENCE, FUNCTION, AND TISSUE
RP   SPECIFICITY.
RX   MEDLINE=20383536; PubMed=10929711;
RA   Du C., Fang M., Li Y., Li L., Wang X.;
RT   "Smac, a mitochondrial protein that promotes cytochrome c-dependent
RT   caspase activation by eliminating IAP inhibition.";
RL   Cell 102:33-42(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA   Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA   Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.
RX   PubMed=10950947;
RA   Srinivasula S.M., Datta P., Fan X.J., Fernandes-Alnemri T., Huang Z.,
RA   Alnemri E.S.;
RT   "Molecular determinants of the caspase-promoting activity of
RT   Smac/DIABLO and its role in the death receptor pathway.";
RL   J. Biol. Chem. 275:36152-36157(2000).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Cerebellum;
RA   Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Suzuki Y.,
RA   Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,
RA   Kawakami B., Nagai K., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Muscle, and Uterus;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 56-239.
RX   MEDLINE=20426096; PubMed=10972280;
RA   Chai J., Du C., Wu J.W., Kyin S., Wang X., Shi Y.;
RT   "Structural and biochemical basis of apoptotic activation by
RT   Smac/DIABLO.";
RL   Nature 406:855-862(2000).
RN   [7]
RP   STRUCTURE BY NMR OF 56-64 IN COMPLEX WITH BIRC4.
RX   MEDLINE=21020961; PubMed=11140637;
RA   Liu Z., Sun C., Olejniczak E.T., Meadows R.P., Betz S.F., Oost T.,
RA   Herrmann J., Wu J.C., Fesik S.W.;
RT   "Structural basis for binding of Smac/DIABLO to the XIAP BIR3
RT   domain.";
RL   Nature 408:1004-1008(2000).
CC   -!- FUNCTION: PROMOTES APOPTOSIS BY ACTIVATING CASPASES IN THE
CC       CYTOCHROME C/APAF-1/CASPASE-9 PATHWAY. ACTS BY OPPOSING THE
CC       INHIBITORY ACTIVITY OF INHIBITOR OF APOPTOSIS PROTEINS (IAP).
CC   -!- SUBUNIT: Homodimer. Interacts with BIRC2, BIRC3, BIRC4/XIAP and
CC       BIRC7.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL BUT RELEASED INTO THE CYTOSOL
CC       WHEN CELLS UNDERGO APOPTOSIS.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9NR28-1; Sequence=Displayed;
CC       Name=2; Synonyms=Diablo-S;
CC         IsoId=Q9NR28-2; Sequence=VSP_004397;
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED WITH HIGHEST EXPRESSION
CC       IN TESTIS. EXPRESSION IS ALSO HIGH IN HEART, LIVER, KIDNEY,
CC       SPLEEN, PROSTATE AND OVARY. LOW IN BRAIN, LUNG, THYMUS AND
CC       PERIPHERAL BLOOD LEUKOCYTES.
CC   -!- DOMAIN: The mature N-terminus mediates interaction with
CC       BIRC4/XIAP.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AF262240; AAF87716.1; -.
DR   EMBL; AK024768; BAB14994.1; -.
DR   EMBL; AF298770; AAG22077.1; -.
DR   EMBL; AK057778; BAB71568.1; -.
DR   EMBL; BC004417; AAH04417.1; -.
DR   PDB; 1FEW; 13-SEP-00.
DR   PDB; 1G3F; 10-JAN-01.
DR   PDB; 1G73; 10-JAN-01.
DR   MIM; 605219; -.
DR   GO; GO:0005739; C:mitochondrion; TAS.
DR   GO; GO:0008635; P:caspase activation via cytochrome c; TAS.
DR   GO; GO:0008625; P:induction of apoptosis via death domain rec. . .; TAS.
DR   GO; GO:0006917; P:induction of apoptosis; TAS.
KW   Transit peptide; Mitochondrion; Apoptosis; Alternative splicing;
KW   3D-structure.
FT   TRANSIT       1     55       MITOCHONDRION.
FT   CHAIN        56    239       SMAC PROTEIN.
FT   SITE         56     60       IAP-BINDING MOTIF (BY SIMILARITY).
FT   VARSPLIC      1     60       MAALKSWLSRSVTSFFRYRQCLCVPVVANFKKRCFSELIRP
FT                                WHKTVTIGFGVTLCAVPIA -> MKSDFYF (in
FT                                isoform 2).
FT                                /FTId=VSP_004397.
FT   CONFLICT     32     32       K -> E (IN REF. 4).
FT   CONFLICT     44     44       K -> R (IN REF. 2).
FT   CONFLICT     62    105       MISSING (IN REF. 4).
FT   CONFLICT    165    165       E -> K (IN REF. 4).
SQ   SEQUENCE   239 AA;  27131 MW;  70C2AE0DC654D031 CRC64;

  Query Match             100.0%;  Score 33;  DB 1;  Length 239;
  Best Local Similarity   100.0%;  Pred. No. 2.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db         56 AVPIAQK 62


RESULT 3
EF2_SULTO
ID   EF2_SULTO      STANDARD;      PRT;   736 AA.
AC   Q975H5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Elongation factor 2 (EF-2).
GN   FUSA OR ST0437.
OS   Sulfolobus tokodaii.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=111955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 10545 / 7;
RX   MEDLINE=21456156; PubMed=11572479;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA   Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA   Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA   Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA   Oshima T., Kikuchi H.;
RT   "Complete genome sequence of an aerobic thermoacidophilic
RT   Crenarchaeon, Sulfolobus tokodaii strain7.";
RL   DNA Res. 8:123-140(2001).
CC   -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC       the nascent protein chain from the A-site to the P-site of the
CC       ribosome.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       EF-G/EF-2 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000982; BAB65426.1; ALT_INIT.
DR   HAMAP; MF_00054; -; 1.
DR   InterPro; IPR004543; aEF-2.
DR   InterPro; IPR000795; EF_GTPbind.
DR   InterPro; IPR000640; EFG_C.
DR   InterPro; IPR005517; EFG_IV.
DR   InterPro; IPR004161; EFTU_D2.
DR   InterPro; IPR005225; Small_GTP.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF03764; EFG_IV; 1.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   TIGRFAMs; TIGR00490; aEF-2; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS00301; EFACTOR_GTP; 1.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   NP_BIND      26     33       GTP (BY SIMILARITY).
FT   NP_BIND      92     96       GTP (BY SIMILARITY).
FT   NP_BIND     146    149       GTP (BY SIMILARITY).
FT   MOD_RES     603    603       DIPHTHAMIDE (BY SIMILARITY).
SQ   SEQUENCE   736 AA;  81777 MW;  226D767F819971ED CRC64;

  Query Match             90.9%;  Score 30;  DB 1;  Length 736;
  Best Local Similarity   85.7%;  Pred. No. 40;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              :||||||
Db        210 SVPIAQK 216


RESULT 4
CERD_CERCA
ID   CERD_CERCA     STANDARD;      PRT;    71 AA.
AC   O17513;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Ceratotoxin D precursor.
GN   CTXD.
OS   Ceratitis capitata (Mediterranean fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Tephritoidea; Tephritidae; Ceratitis.
OX   NCBI_TaxID=7213;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Female accessory gland;
RX   MEDLINE=98231103; PubMed=9569644;
RA   Rosetto M., de Filippis T., Manetti A.G.O., Marchini D., Baldari C.T.,
RA   Dallai R.;
RT   "The genes encoding the antibacterial sex-specific peptides
RT   ceratotoxins are clustered in the genome of the medfly Ceratitis
RT   capitata.";
RL   Insect Biochem. Mol. Biol. 27:1039-1046(1997).
CC   -!- FUNCTION: FEMALE-SPECIFIC PEPTIDES WITH POTENT ACTIVITY AGAINST
CC       GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA. THEY HAVE AS WELL
CC       HEMOLYTIC ACTIVITY. THESE PROTEINS ARE STABLE EVEN AT 100 DEGREES
CC       CELSIUS.
CC   -!- SUBUNIT: HOMOPOLYMER OF FOUR TO SIX SUBUNITS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- SIMILARITY: STRUCTURALLY RELATED TO CECROPINS, DEFENSINS AND
CC       APIADECINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Y15375; CAA75598.1; -.
KW   Insect immunity; Hemolysis; Antibiotic; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24     35       BY SIMILARITY.
FT   PEPTIDE      36     71       CERATOTOXIN D.
SQ   SEQUENCE   71 AA;  7255 MW;  2B4B28E3D2B48516 CRC64;

  Query Match             87.9%;  Score 29;  DB 1;  Length 71;
  Best Local Similarity   85.7%;  Pred. No. 6.8;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||:|
Db         44 AVPIAKK 50


RESULT 5
ARCA_MYCTU
ID   ARCA_MYCTU     STANDARD;      PRT;   402 AA.
AC   O05585;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Arginine deiminase (EC 3.5.3.6) (ADI) (Arginine dihydrolase) (AD).
GN   ARCA OR RV1001 OR MT1030 OR MTCI237.16.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: L-arginine + H(2)O = L-citrulline + NH(3).
CC   -!- PATHWAY: Arginine degradation via arginine deiminase; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: Belongs to the arginine deiminase family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z94752; CAB08144.1; -.
DR   EMBL; AE006986; AAK45280.1; -.
DR   PIR; D70602; D70602.
DR   TIGR; MT1030; -.
DR   TubercuList; Rv1001; -.
DR   HAMAP; MF_00242; -; 1.
DR   InterPro; IPR003876; Arg_deiminase.
DR   Pfam; PF02274; Amidinotransf; 1.
DR   PRINTS; PR01466; ARGDEIMINASE.
DR   TIGRFAMs; TIGR01078; arcA; 1.
KW   Hydrolase; Arginine metabolism; Complete proteome.
SQ   SEQUENCE   402 AA;  43089 MW;  16E5B4BEAA1745D2 CRC64;

  Query Match             87.9%;  Score 29;  DB 1;  Length 402;
  Best Local Similarity   85.7%;  Pred. No. 38;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||||:
Db        257 AVPIAQQ 263


RESULT 6
SYW_PYRAB
ID   SYW_PYRAB      STANDARD;      PRT;   385 AA.
AC   Q9UY11;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase)
DE   (TrpRS).
GN   TRPS OR PYRAB16970 OR PAB1111.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GE5 / Orsay;
RX   PubMed=12622808;
RA   Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA   Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA   Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT   "An integrated analysis of the genome of the hyperthermophilic
RT   archaeon Pyrococcus abyssi.";
RL   Mol. Microbiol. 47:1495-1512(2003).
CC   -!- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
CC       diphosphate + L-tryptophanyl-tRNA(Trp).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AJ248288; CAB50601.1; -.
DR   PIR; C75020; C75020.
DR   HAMAP; MF_00140; -; 1.
DR   InterPro; IPR002305; tRNA-synt_1b.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002306; Trp_tRNA-synt_1b.
DR   Pfam; PF00579; tRNA-synt_1b; 1.
DR   PRINTS; PR01039; TRNASYNTHTRP.
DR   TIGRFAMs; TIGR00233; trpS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
```

```
FT   SITE         82     90       "HIGH" REGION.
FT   SITE        253    257       "KMSKS" REGION.
SQ   SEQUENCE   385 AA;  45100 MW;  4C29D01414976B12 CRC64;

  Query Match             84.8%;  Score 28;  DB 1;  Length 385;
  Best Local Similarity   71.4%;  Pred. No. 62;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:|||:|
Db        165 AIPIAKK 171


RESULT 7
SYW_PYRFU
ID   SYW_PYRFU      STANDARD;      PRT;   385 AA.
AC   Q8U453;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase)
DE   (TrpRS).
GN   TRPS OR PF0241.
OS   Pyrococcus furiosus.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=2261;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA   Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT   "The complete sequence of the Pyrococcus furiosus genome.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
CC       diphosphate + L-tryptophanyl-tRNA(Trp).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE010149; AAL80365.1; -.
DR   HAMAP; MF_00140; -; 1.
DR   InterPro; IPR002305; tRNA-synt_1b.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002306; Trp_tRNA-synt_1b.
DR   Pfam; PF00579; tRNA-synt_1b; 1.
DR   TIGRFAMs; TIGR00233; trpS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
FT   SITE         82     90       "HIGH" REGION.
FT   SITE        253    257       "KMSKS" REGION.
SQ   SEQUENCE   385 AA;  45178 MW;  3A7A628958200CCC CRC64;

  Query Match             84.8%;  Score 28;  DB 1;  Length 385;
  Best Local Similarity   71.4%;  Pred. No. 62;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:|||:|
Db        165 AIPIAKK 171


RESULT 8
SYW_PYRHO
ID   SYW_PYRHO      STANDARD;      PRT;   386 AA.
AC   O59584;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase)
DE   (TrpRS).
GN   TRPS OR PH1921.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
CC       diphosphate + L-tryptophanyl-tRNA(Trp).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000007; BAA31046.1; ALT_INIT.
DR   HAMAP; MF_00140; -; 1.
DR   InterPro; IPR002305; tRNA-synt_1b.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002306; Trp_tRNA-synt_1b.
DR   Pfam; PF00579; tRNA-synt_1b; 1.
DR   PRINTS; PR01039; TRNASYNTHTRP.
DR   TIGRFAMs; TIGR00233; trpS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
FT   SITE         82     90       "HIGH" REGION.
FT   SITE        253    257       "KMSKS" REGION.
SQ   SEQUENCE   386 AA;  45305 MW;  9E3C392F4028B2DD CRC64;

  Query Match             84.8%;  Score 28;  DB 1;  Length 386;
  Best Local Similarity   71.4%;  Pred. No. 62;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:|||:|
Db        165 AIPIAKK 171


RESULT 9
VE2_CRPVK
ID   VE2_CRPVK      STANDARD;      PRT;   390 AA.
AC   P03121;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Probable regulatory protein E2.
GN   E2.
OS   Cottontail rabbit (shope) papillomavirus (strain Kansas) (CRPV).
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=31553;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85166175; PubMed=2984661;
RA   Giri I., Danos O., Yaniv M.;
RT   "Genomic structure of the cottontail rabbit (Shope) papillomavirus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:1580-1584(1985).
CC   -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC       IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
CC       IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC       ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC       WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC       BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC       INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC       REPLICATION.
CC   -!- SUBUNIT: Binds DNA as a dimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02708; -; NOT_ANNOTATED_CDS.
DR   HSSP; P17383; 1DHM.
DR   InterPro; IPR000427; E2_C.
DR   InterPro; IPR001866; E2_N.
DR   Pfam; PF00511; E2_C; 1.
DR   Pfam; PF00508; E2_N; 1.
DR   ProDom; PD000672; E2_C; 1.
DR   ProDom; PD000678; E2_N; 1.
KW   Early protein; Transcription regulation; Activator; DNA-binding;
KW   Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ   SEQUENCE   390 AA;  44024 MW;  8D6B35045E1B4B08 CRC64;

  Query Match           84.8%;  Score 28;  DB 1;  Length 390;
  Best Local Similarity   85.7%;  Pred. No. 63;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||| |||
Db        222 AVPAAQK 228


RESULT 10
BAC2_MOUSE
ID   BAC2_MOUSE     STANDARD;      PRT;   716 AA.
AC   P97303;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transcription regulator protein BACH2 (BTB and CNC homolog 2).
GN   BACH2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=97042438; PubMed=8887638;
RA   Oyake T., Itoh K., Motohashi H., Hayashi N., Hoshino H., Nishizawa M.,
RA   Yamamoto M., Igarashi K.;
RT   "Bach proteins belong to a novel family of BTB-basic leucine zipper
RT   transcription factors that interact with MafK and regulate
RT   transcription through the NF-E2 site.";
RL   Mol. Cell. Biol. 16:6083-6095(1996).
CC   -!- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS AS REPRESSOR OR
CC       ACTIVATOR. BINDS TO MAF RECOGNITION ELEMENTS (MARE). PLAY
CC       IMPORTANT ROLES IN COORDINATING TRANSCRIPTION ACTIVATION AND
CC       REPRESSION BY MAFK.
CC   -!- SUBUNIT: Heterodimer of BACH2 and Maf-related transcription
CC       factors.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: EXPRESSION RESTRICTED TO MONOCYTES AND
CC       NEURONAL CELLS.
CC   -!- SIMILARITY: Belongs to the bZIP family. CNC subfamily.
CC   -!- SIMILARITY: Contains 1 BTB/POZ domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D86604; BAA13138.1; -.
DR   HSSP; P05412; 1FOS.
DR   TRANSFAC; T04792; -.
DR   MGD; MGI:894679; Bach2.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR004827; TF_bZIP.
DR   Pfam; PF00651; BTB; 1.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
DR   SMART; SM00225; BTB; 1.
DR   PROSITE; PS50097; BTB; 1.
DR   PROSITE; PS50217; BZIP; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; Activator; Repressor; DNA-binding;
KW   Nuclear protein.
FT   DOMAIN       37    103       BTB.
FT   DOMAIN      162    168       POLY-GLU.
FT   DNA_BIND    527    542       BASIC MOTIF.
FT   DOMAIN      550    571       LEUCINE-ZIPPER.
SQ   SEQUENCE   716 AA;  78935 MW;  9132B3731AE24333 CRC64;

  Query Match           84.8%;  Score 28;  DB 1;  Length 716;
  Best Local Similarity   57.1%;  Pred. No. 1.1e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:|:|:|
Db        195 AIPVAEK 201


RESULT 11
BAC2_HUMAN
ID   BAC2_HUMAN     STANDARD;      PRT;   841 AA.
AC   Q9BYV9; Q9NTS5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transcription regulator protein BACH2 (BTB and CNC homolog 2).
GN   BACH2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=20404861; PubMed=10949928;
RA   Sasaki S., Ito E., Toki T., Maekawa T., Kanezaki R., Umenai T.,
RA   Muto A., Nagai H., Kinoshita T., Yamamoto M., Inazawa J., Taketo M.M.,
RA   Nakahata T., Igarashi K., Yokoyama M.;
RT   "Cloning and expression of human B cell-specific transcription factor
RT   BACH2 mapped to chromosome 6q15.";
RL   Oncogene 19:3739-3749(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Melo J.V., Vieira S.D., Deininger M.W.N.;
RT   "BACH2 expression in leukaemic cells.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [3]
RP   SEQUENCE OF 1-612 FROM N.A. (ISOFORM 2).
RA   Tromans A.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS AS REPRESSOR OR
CC       ACTIVATOR. BINDS TO MAF RECOGNITION ELEMENTS (MARE). PLAY
CC       IMPORTANT ROLES IN COORDINATING TRANSCRIPTION ACTIVATION AND
CC       REPRESSION BY MAFK (By similarity).
CC   -!- SUBUNIT: Heterodimer of BACH2 and Maf-related transcription
CC       factors (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9BYV9-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9BYV9-2; Sequence=VSP_000582;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: B-cell specific.
CC   -!- SIMILARITY: Belongs to the bZIP family. CNC subfamily.
CC   -!- SIMILARITY: Contains 1 BTB/POZ domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF357835; AAK48898.1; -.
DR   EMBL; AJ271878; CAC28130.1; -.
DR   EMBL; AL121787; CAB87587.1; -.
DR   HSSP; P05412; 1FOS.
DR   TRANSFAC; T04795; -.
DR   Genew; HGNC:14078; BACH2.
DR   MIM; 605394; -.
DR   InterPro; IPR000210; BTB_POZ.
DR   InterPro; IPR004827; TF_bZIP.
DR   Pfam; PF00651; BTB; 1.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
DR   SMART; SM00225; BTB; 1.
DR   PROSITE; PS50097; BTB; 1.
DR   PROSITE; PS50217; BZIP; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; Activator; Repressor; DNA-binding;
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN       37    103       BTB.
FT   DOMAIN      162    169       POLY-GLU.
FT   DNA_BIND    651    666       BASIC MOTIF.
FT   DOMAIN      674    695       LEUCINE-ZIPPER.
FT   VARSPLIC    416    539       Missing (in isoform 2).
FT                                /FTId=VSP_000582.
FT   CONFLICT    291    291       L -> F (IN REF. 1).
SQ   SEQUENCE   841 AA;  92536 MW;  4E926AC325952A93 CRC64;

  Query Match             84.8%;  Score 28;  DB 1;  Length 841;
  Best Local Similarity   57.1%;  Pred. No. 1.3e+02;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:|:|:|
Db        196 AIPVAEK 202


RESULT 12
METH_MYCLE
ID   METH_MYCLE     STANDARD;      PRT;  1206 AA.
AC   Q49775; Q9CC37; Q9S378;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)
DE   (Methionine synthase, vitamin-B12 dependent isozyme) (MS).
GN   METH OR ML1307 OR MLCB2533.04 OR B2126_C1_157.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith D.R., Robison K.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- CATALYTIC ACTIVITY: 5-methyltetrahydrofolate + L-homocysteine =
CC       tetrahydrofolate + L-methionine.
CC   -!- COFACTOR: COBALAMIN (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC   -!- SIMILARITY: BELONGS TO THE VITAMIN-B12 DEPENDENT METHIONINE
CC       SYNTHASE FAMILY.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 873.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00017; AAA17182.1; ALT_FRAME.
DR   EMBL; AL035310; CAA22918.1; ALT_INIT.
DR   EMBL; AL583921; CAC31688.1; -.
DR   PIR; E87072; E87072.
DR   HSSP; P13009; 1BMT.
DR   Leproma; ML1307; -.
DR   InterPro; IPR006158; B12-binding.
DR   InterPro; IPR003759; CoMet_synt_B12.
DR   InterPro; IPR000489; Dhdropt_synt.
DR   InterPro; IPR004223; Met_synt_B12.
DR   InterPro; IPR003726; S_methyl_trans.
DR   Pfam; PF02310; B12-binding; 1.
DR   Pfam; PF02607; B12-binding_2; 1.
DR   Pfam; PF02965; Met_synt_B12; 1.
DR   Pfam; PF00809; Pterin_bind; 1.
DR   Pfam; PF02574; S-methyl_trans; 1.
KW   Transferase; Methyltransferase; Methionine biosynthesis; Vitamin B12;
KW   Cobalt; Complete proteome.
FT   DOMAIN      751    830       COBALAMIN-BINDING (POTENTIAL).
FT   METAL       753    753       COBALT (POTENTIAL).
SQ   SEQUENCE  1206 AA;  132392 MW;  7786CE5307D7CA86 CRC64;

  Query Match             84.8%;  Score 28;  DB 1;  Length 1206;
  Best Local Similarity   71.4%;  Pred. No. 1.9e+02;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |:| |||
Db        341 AIPFAQK 347
```

```
RESULT 13
VGLP_BEV
ID   VGLP_BEV       STANDARD;      PRT;  1581 AA.
AC   P23052;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-NOV-1991 (Rel. 20, Last annotation update)
DE   Peplomer glycoprotein precursor.
GN   P.
OS   Berne virus (BEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Torovirus.
OX   NCBI_TaxID=11156;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Isolate P138/72;
RX   MEDLINE=91020973; PubMed=2219698;
RA   Snijder E.J., den Boon J.A., Spaan W.J.M., Weiss M., Horzinek M.C.;
RT   "Primary structure and post-translational processing of the Berne
RT   virus peplomer protein.";
RL   Virology 178:355-363(1990).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52506; CAA36748.1; -.
DR   PIR; A36759; VGWJBV.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20   1581       PEPLOMER GLYCOPROTEIN.
FT   TRANSMEM   1547   1572       POTENTIAL.
FT   CARBOHYD     25     25       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    384    384       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    494    494       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    574    574       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    935    935       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    969    969       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1267   1267       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1297   1297       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1385   1385       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1389   1389       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1428   1428       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1431   1431       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1438   1438       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1483   1483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1487   1487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1495   1495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1515   1515       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1581 AA;  178332 MW;  00D91B41837AC769 CRC64;

  Query Match             84.8%;  Score 28;  DB 1;  Length 1581;
  Best Local Similarity   83.3%;  Pred. No. 2.5e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              ||:|||
Db        278 VPVAQK 283


RESULT 14
POLG_CX16G
ID   POLG_CX16G     STANDARD;      PRT;  2193 AA.
AC   Q65900;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
DE   (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Core protein
DE   P2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-
DE   linked protein VPG (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C)
DE   (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
OS   Coxsackievirus A16 (strain G-10).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Enterovirus.
OX   NCBI_TaxID=69159;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94303216; PubMed=8030260;
RA   Poyry T., Hyypiae T., Horsnell C., Kinnunen L., Hovi T., Stanway G.;
RT   "Molecular analysis of coxsackievirus A16 reveals a new genetic group
RT   of enteroviruses.";
RL   Virology 202:982-987(1994).
CC   -!- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
CC       MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
CC   -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC       Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC   -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-|-Gly bond in the
CC       poliovirus polyprotein. In other picornavirus reactions Glu may be
CC       substituted for Gln, and Ser or Thr for Gly.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U05876; AAA50478.1; -.
DR   HSSP; P03300; 1POV.
DR   MEROPS; C03.022; -.
DR   MEROPS; C03.UPA; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR000199; Cys_protease_3C.
DR   InterPro; IPR003138; Pico_P1A.
DR   InterPro; IPR000081; Pico_P2A.
DR   InterPro; IPR002527; Pico_P2B.
DR   InterPro; IPR001676; Rhv.
DR   InterPro; IPR000605; RNA_helicase.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR001205; RNA_pol_P3D.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF00548; Cys-protease-3C; 1.
DR   Pfam; PF02226; Pico_P1A; 1.
DR   Pfam; PF00947; Pico_P2A; 1.
DR   Pfam; PF01552; Pico_P2B; 1.
DR   Pfam; PF00073; rhv; 3.
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   Pfam; PF00910; RNA_helicase; 1.
DR   ProDom; PD001125; Cys_protease_3C; 1.
DR   ProDom; PD001306; Pico_P2A; 1.
DR   ProDom; PD001274; Pico_P2B; 1.
DR   SMART; SM00382; AAA; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.
FT   CHAIN         2     69       COAT PROTEIN VP4.
FT   CHAIN        70    323       COAT PROTEIN VP2.
FT   CHAIN       324    565       COAT PROTEIN VP3.
FT   CHAIN       566    862       COAT PROTEIN VP1.
FT   CHAIN       863   1012       CORE PROTEIN P2A.
FT   CHAIN      1013   1111       CORE PROTEIN P2B.
```

```
FT   CHAIN      1112   1440       CORE PROTEIN P2C.
FT   CHAIN      1441   1526       CORE PROTEIN P3A.
FT   CHAIN      1527   1548       GENOME-LINKED PROTEIN VPG.
FT   CHAIN      1549   1731       PICORNAIN 3C.
FT   CHAIN      1732   2193       RNA-DIRECTED RNA POLYMERASE.
FT   LIPID         2      2       MYRISTATE (BY SIMILARITY).
FT   ACT_SITE   1695   1695       PROTEASE (POTENTIAL).
FT   ACT_SITE   1709   1709       PROTEASE (POTENTIAL).
SQ   SEQUENCE   2193 AA;  243209 MW;  04B3BCE572A76E38 CRC64;

  Query Match             84.8%;  Score 28;  DB 1;  Length 2193;
  Best Local Similarity   83.3%;  Pred. No. 3.5e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              :|||||
Db       1105 IPIAQK 1110


RESULT 15
POLG_HE71M
ID   POLG_HE71M     STANDARD;      PRT;  2193 AA.
AC   Q66479;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
DE   (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Core protein
DE   P2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-
DE   linked protein VPG (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C)
DE   (P3C); RNA-directed RNA polymerase P3D (EC 2.7.7.48)].
OS   Human enterovirus 71 (strain 7423/MS/87) (Ev 71).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Enterovirus.
OX   NCBI_TaxID=103922;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96434998; PubMed=8837884;
RA   Brown B.A., Pallansch M.A.;
RT   "Complete nucleotide sequence of enterovirus 71 is distinct from
RT   poliovirus.";
RL   Virus Res. 39:195-206(1995).
CC   -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC       Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC   -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-|-Gly bond in the
CC       poliovirus polyprotein. In other picornavirus reactions Glu may be
CC       substituted for Gln, and Ser or Thr for Gly.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U22522; AAB39969.1; -.
DR   HSSP; P03300; 1POV.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR000199; Cys_protease_3C.
DR   InterPro; IPR003138; Pico_P1A.
DR   InterPro; IPR000081; Pico_P2A.
DR   InterPro; IPR002527; Pico_P2B.
DR   InterPro; IPR001676; Rhv.
DR   InterPro; IPR000605; RNA_helicase.
DR   InterPro; IPR007095; RNA_pol_DS_PS.
DR   InterPro; IPR001205; RNA_pol_P3D.
DR   InterPro; IPR007094; RNA_pol_PSvir.
DR   Pfam; PF00548; Cys-protease-3C; 1.
DR   Pfam; PF02226; Pico_P1A; 1.
DR   Pfam; PF00947; Pico_P2A; 1.
DR   Pfam; PF01552; Pico_P2B; 1.
DR   Pfam; PF00073; rhv; 3.
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   Pfam; PF00910; RNA_helicase; 1.
DR   ProDom; PD001125; Cys_protease_3C; 1.
DR   ProDom; PD001306; Pico_P2A; 1.
DR   ProDom; PD001274; Pico_P2B; 1.
DR   SMART; SM00382; AAA; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.
FT   CHAIN         2     69       COAT PROTEIN VP4.
FT   CHAIN        70    323       COAT PROTEIN VP2.
FT   CHAIN       324    565       COAT PROTEIN VP3.
FT   CHAIN       566    862       COAT PROTEIN VP1.
FT   CHAIN       863   1012       CORE PROTEIN P2A.
FT   CHAIN      1013   1111       CORE PROTEIN P2B.
FT   CHAIN      1112   1440       CORE PROTEIN P2C.
FT   CHAIN      1441   1526       CORE PROTEIN P3A.
FT   CHAIN      1527   1548       GENOME-LINKED PROTEIN VPG.
FT   CHAIN      1549   1731       PICORNAIN 3C.
FT   CHAIN      1732   2193       RNA-DIRECTED RNA POLYMERASE P3D.
FT   LIPID         2      2       MYRISTATE (BY SIMILARITY).
FT   ACT_SITE   1695   1695       PROTEASE (POTENTIAL).
FT   ACT_SITE   1709   1709       PROTEASE (POTENTIAL).
SQ   SEQUENCE   2193 AA;  242656 MW;  35E1B3CFF88A50EF CRC64;

  Query Match             84.8%;  Score 28;  DB 1;  Length 2193;
  Best Local Similarity   83.3%;  Pred. No. 3.5e+02;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              :|||||
Db       1105 IPIAQK 1110


Search completed: September 12, 2003, 11:13:55
Job time : 25 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 12, 2003, 11:10:16 ; Search time 29 Seconds
(without alignments)
10.213 Million cell updates/sec

Title: US-09-939-293A-19_COPY_56_62
Perfect score: 33
Sequence: 1 AVPIAQK 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	100.0	239	3	US-09-479-309-2	Sequence 2, Appli
2	33	100.0	239	4	US-09-627-393-2	Sequence 2, Appli
3	28	84.8	629	4	US-09-252-991A-25963	Sequence 25963, A
4	28	84.8	1190	4	US-09-107-532A-7146	Sequence 7146, Ap
5	27	81.8	326	4	US-09-328-352-8139	Sequence 8139, Ap
6	27	81.8	612	2	US-08-752-307B-11	Sequence 11, Appl
7	27	81.8	612	4	US-09-707-802-11	Sequence 11, Appl
8	27	81.8	612	4	US-09-991-326-11	Sequence 11, Appl
9	27	81.8	1268	4	US-08-506-296B-28	Sequence 28, Appl
10	26	78.8	15	4	US-09-009-953-8	Sequence 8, Appli
11	26	78.8	15	4	US-09-009-953-12	Sequence 12, Appl
12	26	78.8	15	4	US-09-009-953-64	Sequence 64, Appl
13	26	78.8	15	4	US-09-009-953-75	Sequence 75, Appl
14	26	78.8	15	4	US-09-311-784A-414	Sequence 414, App
15	26	78.8	16	4	US-09-009-953-274	Sequence 274, App
16	26	78.8	16	4	US-09-311-784A-413	Sequence 413, App
17	26	78.8	114	4	US-09-107-532A-4832	Sequence 4832, Ap
18	26	78.8	218	4	US-09-130-491-14	Sequence 14, Appl
19	26	78.8	218	4	US-09-489-847-209	Sequence 209, App
20	26	78.8	221	4	US-09-252-991A-18874	Sequence 18874, A
21	26	78.8	282	1	US-07-712-476A-5	Sequence 5, Appli
22	26	78.8	296	1	US-07-712-476A-1	Sequence 1, Appli
23	26	78.8	296	4	US-09-107-532A-6356	Sequence 6356, Ap
24	26	78.8	299	4	US-09-328-352-7407	Sequence 7407, Ap
25	26	78.8	300	6	5340934-6	Patent No. 5340934
26	26	78.8	314	4	US-09-134-253-1	Sequence 1, Appli
27	26	78.8	314	4	US-09-206-576-2	Sequence 2, Appli
28	26	78.8	373	4	US-09-252-991A-24034	
29	26	78.8	457	4	US-09-252-991A-17452	
30	26	78.8	533	4	US-09-388-743-10	
31	26	78.8	559	1	US-08-313-288B-14	
32	26	78.8	667	4	US-09-328-352-5357	
33	26	78.8	947	4	US-09-252-991A-21335	
34	25	75.8	29	3	US-09-143-124-24	
35	25	75.8	148	1	US-07-708-038-4	[illegible], Appli
36	25	75.8	148	1	US-08-127-995-4	Sequence 4, Appli
37	25	75.8	175	4	US-09-252-991A-18834	Sequence 18834, A
38	25	75.8	242	4	US-09-198-452A-182	Sequence 182, App
39	25	75.8	254	4	US-09-252-991A-21540	Sequence 21540, A
40	25	75.8	301	4	US-09-252-991A-33112	Sequence 33112, A
41	25	75.8	302	4	US-09-328-352-6199	Sequence 6199, Ap
42	25	75.8	425	1	US-07-708-038-2	Sequence 2, Appli
43	25	75.8	425	1	US-08-127-995-2	Sequence 2, Appli
44	25	75.8	425	1	US-08-764-343-3	Sequence 3, Appli
45	25	75.8	425	2	US-08-989-925-4	Sequence 4, Appli

ALIGNMENTS

check full seq #6 similar to seq 19? 102(e)?

RESULT 1
US-09-479-309-2
; Sequence 2, Application US/09479309
; Patent No. 6110691
; GENERAL INFORMATION:
; APPLICANT: Wang, Xiaodong
; APPLICANT: Du, Chunying
; TITLE OF INVENTION: Activators of Caspases
; FILE REFERENCE: UTSD0630
; CURRENT APPLICATION NUMBER: US/09/479,309
; CURRENT FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: human
US-09-479-309-2

Query Match 100.0%; Score 33; DB 3; Length 239;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 56 AVPIAQK 62

RESULT 2
US-09-627-393-2
; Sequence 2, Application US/09627393
; Patent No. 6534267
; GENERAL INFORMATION:
; APPLICANT: Wang, Xiaodong
; APPLICANT: Du, Chunying
; TITLE OF INVENTION: Activators of Caspases
; FILE REFERENCE: UTSD0630
; CURRENT APPLICATION NUMBER: US/09/627,393
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/479,309
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 239
; TYPE: PRT
; ORGANISM: human
US-09-627-393-2

Query Match 100.0%; Score 33; DB 4; Length 239;

```
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db         56 AVPIAQK 62


RESULT 3
US-09-252-991A-25963
; Sequence 25963, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 25963
;   LENGTH: 629
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25963

  Query Match             84.8%;  Score 28;  DB 4;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQ 6
              ||||||
Db        200 AVPIAQ 205


RESULT 4
US-09-107-532A-7146
; Sequence 7146, Application US/09107532A
; Patent No. 6583275
;   GENERAL INFORMATION:
;        APPLICANT: Lynn A Doucette-Stamm  and David Bush
;        TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                            ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;        NUMBER OF SEQUENCES: 7310
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;             STREET: 100 Beaver Street
;             CITY: Waltham
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02354
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: CD/ROM ISO9660
;             COMPUTER: PC
;             OPERATING SYSTEM: <Unknown>
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/107,532A
;             FILING DATE: 30-Jun-1998
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/085,598
;             FILING DATE: 14 May 1998
;             APPLICATION NUMBER: 60/051571
;             FILING DATE: July 2, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Ariniello, Pamela Deneke
;             REGISTRATION NUMBER: 40,489
;             REFERENCE/DOCKET NUMBER: GTC-012
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (781)893-5007
;             TELEFAX: (781)893-8277
;   INFORMATION FOR SEQ ID NO: 7146:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1190 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Enterococcus faecium
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION: (B) LOCATION 1...1190
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7146:
US-09-107-532A-7146

  Query Match             84.8%;  Score 28;  DB 4;  Length 1190;
  Best Local Similarity   85.7%;  Pred. No. 3.6e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              ||||| |
Db        202 AVPIASK 208


RESULT 5
US-09-328-352-8139
; Sequence 8139, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8139
;   LENGTH: 326
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-8139

  Query Match             81.8%;  Score 27;  DB 4;  Length 326;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              :|:|||
Db        294 IPVAQK 299


RESULT 6
US-08-752-307B-11
; Sequence 11, Application US/08752307B
; Patent No. 5952171
;  GENERAL INFORMATION:
;    APPLICANT:  McCarthy, Sean A.
;    APPLICANT:  Gearing, David P.
;    APPLICANT:  Levinson, Douglas A.
;    TITLE OF INVENTION:  METHOD FOR IDENTIFYING GENES
;    TITLE OF INVENTION:  ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
```

```
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/752,307B
;       FILING DATE:  19-NOV-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meiklejohn, Ph.D., Anita L.
;       REGISTRATION NUMBER:  35,283
;       REFERENCE/DOCKET NUMBER:  09404/020001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-542-5070
;       TELEFAX:  617-542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  612 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-752-307B-11

  Query Match           81.8%;  Score 27;  DB 2;  Length 612;
  Best Local Similarity   66.7%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              :|:|||
Db        488 IPVAQK 493


RESULT 7
US-09-707-802-11
; Sequence 11, Application US/09707802
; Patent No. 6391586
;   GENERAL INFORMATION:
;       APPLICANT: McCarthy, Sean A.
;                  Gearing, David P.
;                  Levinson, Douglas A.
;       TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
;                           ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
;       NUMBER OF SEQUENCES: 14
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson, P.C.
;            STREET: 225 Franklin Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: US
;            ZIP: 02110-2804
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: Windows95
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/707,802
;            FILING DATE: 07-No. 6391586-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/752,307
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Meiklejohn, Ph.D., Anita L.
;            REGISTRATION NUMBER: 35,283
;            REFERENCE/DOCKET NUMBER: 09404/020001
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 617-542-5070
;            TELEFAX: 617-542-8906
;            TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 11:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 612 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-707-802-11

  Query Match           81.8%;  Score 27;  DB 4;  Length 612;
  Best Local Similarity   66.7%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              :|:|||
Db        488 IPVAQK 493


RESULT 8
US-09-991-326-11
; Sequence 11, Application US/09991326
; Patent No. 6395872
;   GENERAL INFORMATION:
;       APPLICANT: McCarthy, Sean A.
;                  Gearing, David P.
;                  Levinson, Douglas A.
;       TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
;                           ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
;       NUMBER OF SEQUENCES: 14
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson, P.C.
;            STREET: 225 Franklin Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: US
;            ZIP: 02110-2804
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: Windows95
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/991,326
;            FILING DATE: 21-No. 6395872-2001
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/752,307
;            FILING DATE: 19-NOV-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Meiklejohn, Ph.D., Anita L.
;            REGISTRATION NUMBER: 35,283
;            REFERENCE/DOCKET NUMBER: 09404/020002
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 617-542-5070
;            TELEFAX: 617-542-8906
;            TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 11:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 612 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-991-326-11

  Query Match           81.8%;  Score 27;  DB 4;  Length 612;
  Best Local Similarity   66.7%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              :|:|||
Db        488 IPVAQK 493
```

```
RESULT 9
US-08-506-296B-28
; Sequence 28, Application US/08506296B
; Patent No. 6313265
;  GENERAL INFORMATION:
;    APPLICANT:  Phillips, Greg
;    APPLICANT:  Cunningham, Bruce A.
;    APPLICANT:  Crossin, Kathryn L.
;    TITLE OF INVENTION:  NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
;    TITLE OF INVENTION:  CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute
;      STREET:  10550 No. 6313265th Torrey Pines Road, TPC-8
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  U.S.
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/506,296B
;      FILING DATE:  24-JUL-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  TSRI 488.0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 554-2937
;      TELEFAX:  (619) 554-6312
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1268 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-506-296B-28

  Query Match             81.8%;  Score 27;  DB 4;  Length 1268;
  Best Local Similarity   66.7%;  Pred. No. 6.6e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 VPIAQK 7
              :|:|||
Db        488 IPVAQK 493


RESULT 10
US-09-009-953-8
; Sequence 8, Application US/09009953
; Patent No. 6413517
;   GENERAL INFORMATION:
;        APPLICANT: Sette, Alessandro
;        TITLE OF INVENTION: Identification of Broadly
;                            Reactive DR Restricted Epitopes
;        NUMBER OF SEQUENCES: 274
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/009,953
;             FILING DATE: 21-Jan-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/036,713
;             FILING DATE: 23-JAN-1997
;             APPLICATION NUMBER: US 60/037,432
;             FILING DATE: 07-FEB-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Weber, Ellen Lauver
;             REGISTRATION NUMBER: 32,762
;             REFERENCE/DOCKET NUMBER: 018623-011520US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-576-0200
;             TELEFAX: 415-576-0300
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-009-953-8

  Query Match             78.8%;  Score 26;  DB 4;  Length 15;
  Best Local Similarity   71.4%;  Pred. No. 7.9;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:| |
Db          7 AVPLAMK 13


RESULT 11
US-09-009-953-12
; Sequence 12, Application US/09009953
; Patent No. 6413517
;   GENERAL INFORMATION:
;        APPLICANT: Sette, Alessandro
;        TITLE OF INVENTION: Identification of Broadly
;                            Reactive DR Restricted Epitopes
;        NUMBER OF SEQUENCES: 274
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/009,953
;             FILING DATE: 21-Jan-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/036,713
;             FILING DATE: 23-JAN-1997
;             APPLICATION NUMBER: US 60/037,432
;             FILING DATE: 07-FEB-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Weber, Ellen Lauver
;             REGISTRATION NUMBER: 32,762
;             REFERENCE/DOCKET NUMBER: 018623-011520US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-576-0200
;             TELEFAX: 415-576-0300
;             TELEX: <Unknown>
```

```
;   INFORMATION FOR SEQ ID NO: 12:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-009-953-12

  Query Match             78.8%;  Score 26;  DB 4;  Length 15;
  Best Local Similarity   71.4%;  Pred. No. 7.9;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:| |
Db          8 AVPLAMK 14


RESULT 12
US-09-009-953-64
; Sequence 64, Application US/09009953
; Patent No. 6413517
;   GENERAL INFORMATION:
;        APPLICANT: Sette, Alessandro
;        TITLE OF INVENTION: Identification of Broadly
;                            Reactive DR Restricted Epitopes
;        NUMBER OF SEQUENCES: 274
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/009,953
;             FILING DATE: 21-Jan-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/036,713
;             FILING DATE: 23-JAN-1997
;             APPLICATION NUMBER: US 60/037,432
;             FILING DATE: 07-FEB-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Weber, Ellen Lauver
;             REGISTRATION NUMBER: 32,762
;             REFERENCE/DOCKET NUMBER: 018623-011520US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-576-0200
;             TELEFAX: 415-576-0300
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 64:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 64:
US-09-009-953-64

  Query Match             78.8%;  Score 26;  DB 4;  Length 15;
  Best Local Similarity   71.4%;  Pred. No. 7.9;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:| |
Db          7 AVPLAMK 13


RESULT 13
US-09-009-953-75
; Sequence 75, Application US/09009953
; Patent No. 6413517
;   GENERAL INFORMATION:
;        APPLICANT: Sette, Alessandro
;        TITLE OF INVENTION: Identification of Broadly
;                            Reactive DR Restricted Epitopes
;        NUMBER OF SEQUENCES: 274
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/009,953
;             FILING DATE: 21-Jan-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/036,713
;             FILING DATE: 23-JAN-1997
;             APPLICATION NUMBER: US 60/037,432
;             FILING DATE: 07-FEB-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Weber, Ellen Lauver
;             REGISTRATION NUMBER: 32,762
;             REFERENCE/DOCKET NUMBER: 018623-011520US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-576-0200
;             TELEFAX: 415-576-0300
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 75:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 15 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-09-009-953-75

  Query Match             78.8%;  Score 26;  DB 4;  Length 15;
  Best Local Similarity   71.4%;  Pred. No. 7.9;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:| |
Db          8 AVPLAMK 14


RESULT 14
US-09-311-784A-414
; Sequence 414, Application US/09311784A
; Patent No. 6534482
; GENERAL INFORMATION:
;  APPLICANT: Fikes, John D.
;  APPLICANT:  Hermanson, Gary G.
;  APPLICANT:  Sette, Alessandro
;  APPLICANT:  Ishioka, Glenn Y.
;  APPLICANT:  Livingston, Brian
;  APPLICANT:  Chesnut, Robert W.
;  APPLICANT:  Epimmune Inc.
;  TITLE OF INVENTION: Expression Vectors for Stimulating an
```

```
;  TITLE OF INVENTION:  Immune Response and Methods of Using the Same
;  FILE REFERENCE: 39963-20022.01
;  CURRENT APPLICATION NUMBER: US/09/311,784A
;  CURRENT FILING DATE:  1999-05-13
;  PRIOR APPLICATION NUMBER: US 60/085,751
;  PRIOR FILING DATE: 1998-05-15
;  NUMBER OF SEQ ID NOS: 463
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 414
;   LENGTH: 15
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Pf SSP2 62 (peptide 1188.34)
US-09-311-784A-414

  Query Match             78.8%;  Score 26;  DB 4;  Length 15;
  Best Local Similarity   71.4%;  Pred. No. 7.9;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:| |
Db          7 AVPLAMK 13


RESULT 15
US-09-009-953-274
; Sequence 274, Application US/09009953
; Patent No. 6413517
;   GENERAL INFORMATION:
;        APPLICANT: Sette, Alessandro
;        TITLE OF INVENTION: Identification of Broadly
;                            Reactive DR Restricted Epitopes
;        NUMBER OF SEQUENCES: 274
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/009,953
;             FILING DATE: 21-Jan-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/036,713
;             FILING DATE: 23-JAN-1997
;             APPLICATION NUMBER: US 60/037,432
;             FILING DATE: 07-FEB-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Weber, Ellen Lauver
;             REGISTRATION NUMBER: 32,762
;             REFERENCE/DOCKET NUMBER: 018623-011520US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-576-0200
;             TELEFAX: 415-576-0300
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 274:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 16 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 274:
US-09-009-953-274

  Query Match             78.8%;  Score 26;  DB 4;  Length 16;
  Best Local Similarity   71.4%;  Pred. No. 8.5;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||:| |
Db          8 AVPLAMK 14


Search completed: September 12, 2003, 11:16:12
Job time : 30 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 12, 2003, 11:13:31 ; Search time 26 Seconds
(without alignments)
39.284 Million cell updates/sec

Title: US-09-939-293A-19_COPY_56_62
Perfect score: 33
Sequence: 1 AVPIAQK 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	100.0	7	10	US-09-939-293-6	Sequence 6, Appli
2	33	100.0	7	10	US-09-965-967-8	Sequence 8, Appli
3	33	100.0	7	12	US-10-293-371-1	Sequence 1, Appli
4	33	100.0	7	12	US-10-293-371-24	Sequence 24, Appl
5	33	100.0	7	12	US-10-293-371-45	Sequence 45, Appl
6	33	100.0	7	14	US-10-068-569-12	Sequence 12, Appl
7	33	100.0	10	10	US-09-965-967-18	Sequence 18, Appl
8	33	100.0	13	10	US-09-965-967-25	Sequence 25, Appl
9	33	100.0	15	14	US-10-068-569-8	Sequence 8, Appli
10	33	100.0	15	15	US-10-197-634-8	Sequence 8, Appli
11	33	100.0	30	10	US-09-939-293-7	Sequence 7, Appli
12	33	100.0	35	10	US-09-939-293-11	Sequence 11, Appl
13	33	100.0	39	10	US-09-939-293-8	Sequence 8, Appli
14	33	100.0	40	10	US-09-939-293-2	Sequence 2, Appli
15	33	100.0	84	10	US-09-798-116-9	Sequence 9, Appli
16	33	100.0	202	10	US-09-798-116-7	Sequence 7, Appli
17	33	100.0	227	9	US-09-925-297-591	Sequence 591, App
18	33	100.0	237	10	US-09-798-116-2	Sequence 2, Appli
19	33	100.0	237	10	US-09-798-116-4	Sequence 4, Appli
20	33	100.0	239	12	US-10-141-618-14	Sequence 14, Appl
21	33	100.0	239	15	US-10-153-668-348	Sequence 348, App
22	29	87.9	409	11	US-09-252-088-44	Sequence 44, Appl
23	29	87.9	434	11	US-09-252-088-39	Sequence 39, Appl
24	29	87.9	532	15	US-10-156-761-10999	Sequence 10999, A
25	28	84.8	874	15	US-10-163-214-13	Sequence 13, Appl
26	28	84.8	3564	15	US-10-156-761-7964	Sequence 7964, Ap
27	27	81.8	129	9	US-09-734-569-94	Sequence 94, Appl
28	27	81.8	400	10	US-09-738-626-4488	Sequence 4488, Ap
29	27	81.8	475	9	US-09-734-569-166	Sequence 166, App
30	27	81.8	729	15	US-10-156-761-7902	Sequence 7902, Ap
31	26	78.8	16	10	US-09-894-018-274	Sequence 274, App
32	26	78.8	20	12	US-10-079-167-95	Sequence 95, Appl
33	26	78.8	205	10	US-09-738-626-4863	Sequence 4863, Ap
34	26	78.8	218	11	US-09-978-418-16	Sequence 16, Appl
35	26	78.8	218	12	US-10-365-227-15	Sequence 15, Appl
36	26	78.8	218	14	US-10-105-929-14	Sequence 14, Appl
37	26	78.8	276	10	US-09-894-018-141	Sequence 141, App
38	26	78.8	287	12	US-10-301-822-147	Sequence 147, App
39	26	78.8	287	15	US-10-097-340-231	Sequence 231, App
40	26	78.8	287	15	US-10-171-311-176	Sequence 176, App
41	26	78.8	287	15	US-10-205-823-307	Sequence 307, App
42	26	78.8	287	15	US-10-177-293-340	Sequence 340, App
43	26	78.8	300	9	US-09-919-770-2	Sequence 2, Appli
44	26	78.8	300	12	US-10-301-822-145	Sequence 145, App
45	26	78.8	300	15	US-10-097-340-229	Sequence 229, App

ALIGNMENTS

RESULT 1
US-09-939-293-6
; Sequence 6, Application US/09939293
; Patent No. US20020132786A1
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
; TITLE OF INVENTION: AND METHODS OF USING THE SAME
; FILE REFERENCE: 480140.465
; CURRENT APPLICATION NUMBER: US/09/939,293
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-939-293-6

Query Match 100.0%; Score 33; DB 10; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.8e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7
```

RESULT 2
US-09-965-967-8
; Sequence 8, Application US/09965967
; Patent No. US20020177557A1
; GENERAL INFORMATION:
; APPLICANT: Shi, Yigong
; TITLE OF INVENTION: Compositions And Methods For Regulating Apoptosis
; FILE REFERENCE: PU-0031 (01-1739-1)
; CURRENT APPLICATION NUMBER: US/09/965,967
; CURRENT FILING DATE: 2001-09-28

```
;  PRIOR APPLICATION NUMBER: 60/236,574
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: 60/256,830
;  PRIOR FILING DATE: 2000-12-20
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-965-967-8

  Query Match           100.0%;  Score 33;  DB 10;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 4.8e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 3
US-10-293-371-1
; Sequence 1, Application US/10293371
; Publication No. US20030157522A1
; GENERAL INFORMATION:
;  APPLICANT: BOUDREAULT, ALAIN
;  APPLICANT:  KORNELUK, ROBERT G.
;  APPLICANT:  LACASSE, ERIC
;  APPLICANT:  LISTON, PETER
;  TITLE OF INVENTION: Methods and Reagents for Peptide-Bir
;  TITLE OF INVENTION:  Interaction Screens
;  FILE REFERENCE: 07891/030002
;  CURRENT APPLICATION NUMBER: US/10/293,371
;  CURRENT FILING DATE:  2003-04-08
;  PRIOR APPLICATION NUMBER: US 60/370,934
;  PRIOR FILING DATE: 2002-04-08
;  PRIOR APPLICATION NUMBER: US 60/332,300
;  PRIOR FILING DATE: 2001-11-09
;  NUMBER OF SEQ ID NOS: 85
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-293-371-1

  Query Match           100.0%;  Score 33;  DB 12;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 4.8e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 4
US-10-293-371-24
; Sequence 24, Application US/10293371
; Publication No. US20030157522A1
; GENERAL INFORMATION:
;  APPLICANT: BOUDREAULT, ALAIN
;  APPLICANT:  KORNELUK, ROBERT G.
;  APPLICANT:  LACASSE, ERIC
;  APPLICANT:  LISTON, PETER
;  TITLE OF INVENTION: Methods and Reagents for Peptide-Bir
;  TITLE OF INVENTION:  Interaction Screens
;  FILE REFERENCE: 07891/030002
;  CURRENT APPLICATION NUMBER: US/10/293,371
;  CURRENT FILING DATE:  2003-04-08
;  PRIOR APPLICATION NUMBER: US 60/370,934
;  PRIOR FILING DATE: 2002-04-08
;  PRIOR APPLICATION NUMBER: US 60/332,300
;  PRIOR FILING DATE: 2001-11-09
;  NUMBER OF SEQ ID NOS: 85
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-293-371-24

  Query Match           100.0%;  Score 33;  DB 12;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 4.8e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 5
US-10-293-371-45
; Sequence 45, Application US/10293371
; Publication No. US20030157522A1
; GENERAL INFORMATION:
;  APPLICANT: BOUDREAULT, ALAIN
;  APPLICANT:  KORNELUK, ROBERT G.
;  APPLICANT:  LACASSE, ERIC
;  APPLICANT:  LISTON, PETER
;  TITLE OF INVENTION: Methods and Reagents for Peptide-Bir
;  TITLE OF INVENTION:  Interaction Screens
;  FILE REFERENCE: 07891/030002
;  CURRENT APPLICATION NUMBER: US/10/293,371
;  CURRENT FILING DATE:  2003-04-08
;  PRIOR APPLICATION NUMBER: US 60/370,934
;  PRIOR FILING DATE: 2002-04-08
;  PRIOR APPLICATION NUMBER: US 60/332,300
;  PRIOR FILING DATE: 2001-11-09
;  NUMBER OF SEQ ID NOS: 85
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-10-293-371-45

  Query Match           100.0%;  Score 33;  DB 12;  Length 7;
  Best Local Similarity  100.0%;  Pred. No. 4.8e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 6
US-10-068-569-12
; Sequence 12, Application US/10068569
; Publication No. US20020160975A1
; GENERAL INFORMATION:
;  APPLICANT: Srinivasula, Srinivasa M.
;  APPLICANT:  Fernandes-Alnemri, Teresa
;  APPLICANT:  Alnemri, Emad S.
;  TITLE OF INVENTION: A CONSERVED XIAP-INTERACTION MOTIF IN
;  TITLE OF INVENTION:  CASPASE-9 AND SMAC/DIABLO FOR MEDIATING APOPTOSIS
;  FILE REFERENCE: 480140.475
;  CURRENT APPLICATION NUMBER: US/10/068,569
;  CURRENT FILING DATE:  2002-02-06
;  NUMBER OF SEQ ID NOS: 28
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-068-569-12

  Query Match             100.0%;  Score 33;  DB 14;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 4.8e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 7
US-09-965-967-18
; Sequence 18, Application US/09965967
; Patent No. US20020177557A1
; GENERAL INFORMATION:
;  APPLICANT: Shi, Yigong
;  TITLE OF INVENTION: Compositions And Methods For Regulating Apoptosis
;  FILE REFERENCE: PU-0031 (01-1739-1)
;  CURRENT APPLICATION NUMBER: US/09/965,967
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: 60/236,574
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: 60/256,830
;  PRIOR FILING DATE: 2000-12-20
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-965-967-18

  Query Match             100.0%;  Score 33;  DB 10;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 8
US-09-965-967-25
; Sequence 25, Application US/09965967
; Patent No. US20020177557A1
; GENERAL INFORMATION:
;  APPLICANT: Shi, Yigong
;  TITLE OF INVENTION: Compositions And Methods For Regulating Apoptosis
;  FILE REFERENCE: PU-0031 (01-1739-1)
;  CURRENT APPLICATION NUMBER: US/09/965,967
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: 60/236,574
;  PRIOR FILING DATE: 2000-09-29
;  PRIOR APPLICATION NUMBER: 60/256,830
;  PRIOR FILING DATE: 2000-12-20
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 13
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-965-967-25

  Query Match             100.0%;  Score 33;  DB 10;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 0.57;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          4 AVPIAQK 10


RESULT 9
US-10-068-569-8
; Sequence 8, Application US/10068569
; Publication No. US20020160975A1
; GENERAL INFORMATION:
;  APPLICANT: Srinivasula, Srinivasa M.
;  APPLICANT:  Fernandes-Alnemri, Teresa
;  APPLICANT:  Alnemri, Emad S.
;  TITLE OF INVENTION: A CONSERVED XIAP-INTERACTION MOTIF IN
;  TITLE OF INVENTION:  CASPASE-9 AND SMAC/DIABLO FOR MEDIATING APOPTOSIS
;  FILE REFERENCE: 480140.475
;  CURRENT APPLICATION NUMBER: US/10/068,569
;  CURRENT FILING DATE:  2002-02-06
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 15
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-068-569-8

  Query Match             100.0%;  Score 33;  DB 14;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.66;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 10
US-10-197-634-8
; Sequence 8, Application US/10197634
; Publication No. US20030073629A1
; GENERAL INFORMATION:
;  APPLICANT: Alnemri, Emad S.
;  TITLE OF INVENTION: OMI AND DOMAINS THEREOF THAT DISRUPT
;  TITLE OF INVENTION:  IAP-CASPASE INTERACTION
;  FILE REFERENCE: 480140.479
;  CURRENT APPLICATION NUMBER: US/10/197,634
;  CURRENT FILING DATE:  2002-07-15
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 15
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-197-634-8

  Query Match             100.0%;  Score 33;  DB 15;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.66;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 11
US-09-939-293-7
; Sequence 7, Application US/09939293
; Patent No. US20020132786A1
; GENERAL INFORMATION:
;  APPLICANT: Alnemri, Emad S.
;  TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
;  TITLE OF INVENTION:  AND METHODS OF USING THE SAME
;  FILE REFERENCE: 480140.465
;  CURRENT APPLICATION NUMBER: US/09/939,293
```

```
;  CURRENT FILING DATE:  2001-08-24
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 30
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-939-293-7

  Query Match             100.0%;  Score 33;  DB 10;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 1.4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 12
US-09-939-293-11
; Sequence 11, Application US/09939293
; Patent No. US20020132786A1
; GENERAL INFORMATION:
;  APPLICANT: Alnemri, Emad S.
;  TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
;  TITLE OF INVENTION:  AND METHODS OF USING THE SAME
;  FILE REFERENCE: 480140.465
;  CURRENT APPLICATION NUMBER: US/09/939,293
;  CURRENT FILING DATE:  2001-08-24
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 35
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-939-293-11

  Query Match             100.0%;  Score 33;  DB 10;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 13
US-09-939-293-8
; Sequence 8, Application US/09939293
; Patent No. US20020132786A1
; GENERAL INFORMATION:
;  APPLICANT: Alnemri, Emad S.
;  TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
;  TITLE OF INVENTION:  AND METHODS OF USING THE SAME
;  FILE REFERENCE: 480140.465
;  CURRENT APPLICATION NUMBER: US/09/939,293
;  CURRENT FILING DATE:  2001-08-24
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 39
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-939-293-8

  Query Match             100.0%;  Score 33;  DB 10;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 14
US-09-939-293-2
; Sequence 2, Application US/09939293
; Patent No. US20020132786A1
; GENERAL INFORMATION:
;  APPLICANT: Alnemri, Emad S.
;  TITLE OF INVENTION: AN IAP PEPTIDE OR POLYPEPTIDE
;  TITLE OF INVENTION:  AND METHODS OF USING THE SAME
;  FILE REFERENCE: 480140.465
;  CURRENT APPLICATION NUMBER: US/09/939,293
;  CURRENT FILING DATE:  2001-08-24
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 40
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-939-293-2

  Query Match             100.0%;  Score 33;  DB 10;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 15
US-09-798-116-9
; Sequence 9, Application US/09798116
; Patent No. US20020110851A1
; GENERAL INFORMATION:
;  APPLICANT: Verhagen, Anne Marie
;  APPLICANT:  Ekert, Paul
;  APPLICANT:  Vaux, David
;  TITLE OF INVENTION: No. US20020110851A1el Polypeptides, Modulatory Agents Therefor
;  FILE REFERENCE: 10338-004US
;  CURRENT APPLICATION NUMBER: US/09/798,116
;  CURRENT FILING DATE:  2001-03-02
;  PRIOR APPLICATION NUMBER: AU PQ5995/00
;  PRIOR FILING DATE: 2000-03-02
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 84
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-09-798-116-9

  Query Match             100.0%;  Score 33;  DB 10;  Length 84;
  Best Local Similarity   100.0%;  Pred. No. 3.9;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db         54 AVPIAQK 60


Search completed: September 12, 2003, 11:17:31
Job time : 27 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 12, 2003, 10:56:11 ; Search time 83 Seconds
(without alignments)
13.387 Million cell updates/sec

Title: US-09-939-293A-19_COPY_56_62
Perfect score: 33
Sequence: 1 AVPIAQK 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	100.0	7	23	ABB76213	Human smac (DIABLO
2	33	100.0	7	23	AAU78434	Inhibitor of apopt
3	33	100.0	7	23	AAU78487	Smac-7 AV peptoid.
4	33	100.0	8	23	ABB76212	Human smac (DIABLO
5	33	100.0	9	23	ABB76209	Human smac (DIABLO
6	33	100.0	9	23	ABB76229	Human smac (DIABLO
7	33	100.0	10	23	ABB76228	Fluorosceinated sm
8	33	100.0	15	24	ABP71314	Human Smac protein
9	33	100.0	20	23	ABB76208	Human smac (DIABLO
10	33	100.0	30	23	AAU78435	Inhibitor of apopt
11	33	100.0	35	23	AAU78439	Inhibitor of apopt
12	33	100.0	39	23	AAU78436	Inhibitor of apopt
13	33	100.0	40	23	AAU78430	Inhibitor of apopt
14	33	100.0	84	24	ABG72303	Rat partial sequen
15	33	100.0	202	24	ABG72302	Human partial sequ
16	33	100.0	227	21	AAB54139	Human pancreatic c
17	33	100.0	237	24	ABG72301	Mouse pro-apoptoti
18	33	100.0	239	21	AAB26210	Human caspase acti
19	33	100.0	239	23	AAU78447	Inhibitor of apopt
20	33	100.0	239	24	ABP72164	Human DIABLO/Smac.
21	33	100.0	239	24	ABB82743	Human Smac polypep
22	31	93.9	945	22	ABB63213	Drosophila melanog
23	30	90.9	66	21	AAG00964	Human secreted pro
24	30	90.9	567	22	AAM40580	Human polypeptide
25	30	90.9	578	22	AAB94675	Human protein sequ
26	30	90.9	583	22	AAM38794	Human polypeptide
27	30	90.9	586	22	AAU23484	Novel human enzyme
28	29	87.9	9	23	ABB76218	Human smac (DIABLO
29	29	87.9	9	23	ABB76221	Human smac (DIABLO
30	29	87.9	9	23	ABB76222	Human smac (DIABLO
31	29	87.9	9	23	ABB76225	Human smac (DIABLO
32	29	87.9	9	23	ABB76226	Human smac (DIABLO
33	29	87.9	9	23	ABB76227	Human smac (DIABLO
34	29	87.9	409	20	AAY27370	Group B Streptococ
35	29	87.9	432	23	ABP29682	Streptococcus poly
36	29	87.9	434	20	AAY27367	Group B Streptococ
37	29	87.9	434	23	ABP25888	Streptococcus poly
38	29	87.9	437	22	AAB96580	Putative P. abyssi
39	29	87.9	4694	22	ABG19817	Novel human diagno
40	28	84.8	6	23	ABB76214	Human smac (DIABLO
41	28	84.8	6	23	AAU78486	Smac-6 AV peptoid.
42	28	84.8	9	23	ABB76211	Human smac (DIABLO
43	28	84.8	9	23	ABB76224	Human smac (DIABLO
44	28	84.8	51	22	ABG28508	Novel human diagno
45	28	84.8	51	24	ABR47856	Human secreted pro

need more hits

ALIGNMENTS

RESULT 1
ABB76213
ID ABB76213 standard; Peptide; 7 AA.
XX
AC ABB76213;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human smac (DIABLO) derived peptide.
XX
KW DIABLO; smac; inhibitor of apoptosis protein; IAP; apoptosis;
KW human; cancer; cytostatic; mutant; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 7
FT /note= "optional C-terminal protecting group,
FT e.g. C-terminal amide"
XX
PN WO200230959-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US32121.
XX
PR 13-OCT-2000; 2000US-0687549.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Fesik SW, Meadows RP, Betz SP, Liu Z, Olejniczak ET, Sun C;
XX

DR WPI; 2002-444169/47.
XX
PT Novel peptide derived from wild-type human second mitochondria derived
PT activator of caspase protein useful for identifying candidate
PT substances to kill cancerous cells -
XX
PS Claim 5; Page 7; 26pp; English.
XX
CC The present sequence is a peptide derived from wild-type human
CC second mitochondria derived activator of caspase (smac), also known
CC as direct inhibitor of apoptosis binding protein with low pI
CC (DIABLO). The peptide is one of 12 claimed smac (DIABLO)-derived
CC peptides (see ABB76208-19) which bind to the Bir2 and Bir3 domain
CC of XIAP, an inhibitor of apoptosis protein (IAP) family member.
CC Kd values for Bir-3 and Bir-2 are 0.70 +/- 0.09 uM and 9.4 +/- 0.6
CC uM, respectively, for the present (C-terminally amidated) peptide,
CC compared with 0.42 +/- 0.02 uM and 2.3 +/- 0.3 uM, respectively,
CC for full-length smac. Modification of the N-terminal alanine
CC destroys binding affinity to XIAP, and mutation of the valine,
CC proline or isoleucine also causes some loss of binding. Amino
CC acids C-terminal to the isoleucine are not important for binding.
CC The claimed peptides can be used to identify candidate substances
CC which induce or promote apoptosis in cells. The assay involves
CC determination of the ability of candidate compounds to disrupt the
CC binding interaction between a smac (DIABLO) peptide and an IAP
CC family member.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 23; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 2
AAU78434
ID AAU78434 standard; Peptide; 7 AA.
XX
AC AAU78434;
XX
DT 18-JUN-2002 (first entry)
XX
DE Inhibitor of apoptosis (IAP) protein Smac, mutant Smac-N7.
XX
KW Human; inhibitor of apoptosis; IAP; Smac; apoptosis; BID; BIR1; BIR2;
KW Bcl2 interacting domain; caspase; BIR domain; BIR3; gene therapy;
KW neoplastic cell; mutant; tumour.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200216418-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26492.
XX
PR 24-AUG-2000; 2000US-227735P.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Alnemri ES;
XX
DR WPI; 2002-304115/34.
XX
PT Novel Smac peptides and polynucleotides encoding the peptides, useful
PT for stimulating apoptosis in neoplastic or tumour cell which
PT overexpresses inhibitor of caspase, and for identifying apoptosis
PT modulating compounds -
XX
PS Example 3; Fig 7; 78pp; English.
XX
CC The invention relates to an isolated Smac peptide or polypeptide (I)
CC and an isolated nucleic acid (II) encoding (I). Also described is a
CC method of identifying a compound that inhibits apoptosis, comprising:
CC (a) separately contacting several cell populations expressing a
CC cytosolic Smac (a Smac isoform that begins with MKSDFYF sequence,
CC replacing the mitochondrial targeting sequence (residues 1-55 of (I)),
CC and residues 56-60 of (I)) and an inhibitor of BID (Bcl2 interacting
CC domain) with a compound to be tested for apoptotic inhibiting activity;
CC (b) incubating the cell populations with a direct stimulus of the cell
CC death pathway; and (c) measuring the specific apoptotic activity of the
CC cell populations, where inhibition of the specific apoptotic activity is
CC indicative that the compound is an inhibitor of apoptosis. (I) and (II)
CC are useful for inducing apoptosis in a cell. The Smac polypeptide and
CC polynucleotide are useful for stimulating apoptosis in a neoplastic or
CC tumour cell which overexpresses an inhibitor of caspase, where the
CC inhibitor inhibits activation or activity of caspase-3, caspase-7 or
CC caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
CC (I) is useful for identifying an inhibitor or enhancer of a caspase-
CC mediated apoptosis which involves contacting a cell transformed or
CC transfected with a vector expressing (I) with a candidate inhibitor or
CC candidate enhancer; and detecting cell viability, where an increase in
CC cell viability indicates the presence of an inhibitor and a decrease in
CC cell viability indicates the presence of an enhancer. Optionally, the
CC method involves detecting the presence of large and small caspase
CC subunits after contacting cell transformed with the vector expressing
CC (I), with the candidate compound. A decrease in processing indicates the
CC presence of an inhibitor and an increase in the processing indicates the
CC presence of an enhancer. Preferably, the large and small subunits of
CC caspase-3, caspase-7 or caspase-9 are detected. (I) is also useful for
CC identifying a compound that inhibits Smac binding to Smac-binding
CC molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
CC or a full-length IAP). (II) is useful in gene therapy techniques. The
CC present sequence represents the amino acid sequence of Smac mutant
CC Smac-N7.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 23; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 3
AAU78487
ID AAU78487 standard; Peptide; 7 AA.
XX
AC AAU78487;
XX
DT 18-JUN-2002 (first entry)
XX
DE Smac-7 AV peptoid.
XX
KW Apoptosis; cytostatic; apoptotic; AV peptoid; melanoma; lymphoma;
KW Inhibitor of Apoptosis protein; IAP; procaspase-3; tumour cell;
KW breast cancer; prostate cancer; lung cancer; pancreatic cancer; smac-7;
KW gastric cancer; colon cancer; ovarian cancer; renal cancer; hepatoma;
KW sarcoma; smac; second mitochondria-derived activator of caspases.
XX
OS Synthetic.
XX
PN WO200216402-A2.
XX
PD 28-FEB-2002.
XX
PF 23-AUG-2001; 2001WO-US41869.
XX

PR 23-AUG-2000; 2000US-0645075.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Wang X, Du C;
XX
DR WPI; 2002-280909/32.
XX
PT Composition for enhancing the apoptosis of pathogenic cells,
PT particularly tumour cells, e.g. breast cancer, prostate cancer, lung
PT cancer, colon cancer, ovarian cancer or sarcoma, comprises apoptotic
PT compounds -
XX
PS Example 9; Page 28; 40pp; English.
XX
CC This invention relates to a method for induction of apoptosis in
CC pathogenic cells. The method comprises a novel pharmaceutical
CC composition which comprises an AV peptoid in dosage form and a
CC pharmaceutical carrier, where the AV peptoid comprises a peptide that
CC interacts with or inhibits the activity of an Inhibitor of Apoptosis
CC protein (IAP) as measured by IAP binding, procaspase-3 activation or
CC promotion of apoptosis. The peptoides of the invention are used to
CC inhibit an inhibitor of Apoptosis protein (IAP). Compositions containing
CC these peptoids are useful for enhancing the apoptosis of pathogenic
CC cells, particularly tumour cells, e.g. breast cancer, prostate cancer,
CC lung cancer, pancreatic cancer, gastric cancer, colon cancer, ovarian
CC cancer, renal cancer, hepatoma, melanoma, lymphoma or sarcoma. The
CC composition is particularly useful for promoting cell death. The
CC present sequence represents an AV peptoid (smac-7) used to inhibit
CC second mitochondria-derived activator of caspases (smac) using the
CC method of the invention. Smac interacts with and eliminates the activity
CC of a number of IAP's and as such inhibiting its activity allows the
CC induction of apoptosis.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 33; DB 23; Length 7;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 4
ABB76212
ID ABB76212 standard; Peptide; 8 AA.
XX
AC ABB76212;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human smac (DIABLO) derived peptide.
XX
KW DIABLO; smac; inhibitor of apoptosis protein; IAP; apoptosis;
KW human; cancer; cytostatic; mutant; mutein.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 8
FT /note= "optional C-terminal protecting group,
FT e.g. C-terminal amide"
XX
PN WO200230959-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US32121.
XX
PR 13-OCT-2000; 2000US-0687549.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Fesik SW, Meadows RP, Betz SP, Liu Z, Olejniczak ET, Sun C;
XX
DR WPI; 2002-444169/47.
XX
PT Novel peptide derived from wild-type human second mitochondria derived
PT activator of caspase protein useful for identifying candidate
PT substances to kill cancerous cells -
XX
PS Claim 5; Page 7; 26pp; English.
XX
CC The present sequence is a peptide derived from wild-type human
CC second mitochondria derived activator of caspase (smac), also known
CC as direct inhibitor of apoptosis binding protein with low pI
CC (DIABLO). The peptide is one of 12 claimed smac (DIABLO)-derived
CC peptides (see ABB76208-19) which bind to the Bir2 and Bir3 domain
CC of XIAP, an inhibitor of apoptosis protein (IAP) family member.
CC Kd values for Bir-3 and Bir-2 are 0.80 +/- 0.03 uM and 13 +/- 0.3
CC uM, respectively, for the present (C-terminally amidated) peptide,
CC compared with 0.42 +/- 0.02 uM and 2.3 +/- 0.3 uM, respectively,
CC for full-length smac. Modification of the N-terminal alanine
CC destroys binding affinity to XIAP, and mutation of the valine,
CC proline or isoleucine also causes some loss of binding. Amino
CC acids C-terminal to the isoleucine are not important for binding.
CC The claimed peptides can be used to identify candidate substances
CC which induce or promote apoptosis in cells. The assay involves
CC determination of the ability of candidate compounds to disrupt the
CC binding interaction between a smac (DIABLO) peptide and an IAP
CC family member.
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 33; DB 23; Length 8;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 5
ABB76209
ID ABB76209 standard; Peptide; 9 AA.
XX
AC ABB76209;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human smac (DIABLO) derived peptide.
XX
KW DIABLO; smac; inhibitor of apoptosis protein; IAP; apoptosis;
KW human; cancer; cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 9
FT /note= "optional C-terminal protecting group"
XX
PN WO200230959-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US32121.
XX
PR 13-OCT-2000; 2000US-0687549.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Fesik SW, Meadows RP, Betz SP, Liu Z, Olejniczak ET, Sun C;
XX

DR WPI; 2002-444169/47.
XX
PT Novel peptide derived from wild-type human second mitochondria derived
PT activator of caspase protein useful for identifying candidate
PT substances to kill cancerous cells -
XX
PS Claim 5; Page 7; 26pp; English.
XX
CC The present sequence is a peptide derived from wild-type human
CC second mitochondria derived activator of caspase (smac), also known
CC as direct inhibitor of apoptosis binding protein with low pI
CC (DIABLO). The peptide is one of 12 claimed smac (DIABLO)-derived
CC peptides (see ABB76208-19) which bind to the Bir2 and Bir3 domain
CC of XIAP, an inhibitor of apoptosis protein (IAP) family member.
CC Kd values for Bir-3 and Bir-2 are 0.43 +/- 0.06 uM and 6.0 +/- 0.9
CC uM, respectively, for the present peptide, compared with 0.42 +/-
CC 0.02 uM and 2.3 +/- 0.3 uM, respectively, for full-length smac.
CC Modification of the N-terminal alanine destroys binding affinity to
CC XIAP, and mutation of the valine, proline or isoleucine also causes
CC some loss of binding. Amino acids C-terminal to the isoleucine are
CC not important for binding. The claimed peptides can be used to
CC identify candidate substances which induce or promote apoptosis in
CC cells. The assay involves determination of the ability of
CC candidate compounds to disrupt the binding interaction between a
CC smac (DIABLO) peptide and an IAP family member.
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 33; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 6
ABB76229
ID ABB76229 standard; Peptide; 9 AA.
XX
AC ABB76229;
XX
DT 09-AUG-2002 (first entry)
XX
DE Human smac (DIABLO) derived peptide.
XX
KW DIABLO; smac; inhibitor of apoptosis protein; IAP; apoptosis;
KW human; cancer; cytostatic; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "N-terminal acetyl"
FT Modified-site 9
FT /note= "optional C-terminal protecting group"
XX
PN WO200230959-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US32121.
XX
PR 13-OCT-2000; 2000US-0687549.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Fesik SW, Meadows RP, Betz SP, Liu Z, Olejniczak ET, Sun C;
XX
DR WPI; 2002-444169/47.
XX
PT Novel peptide derived from wild-type human second mitochondria derived
PT activator of caspase protein useful for identifying candidate
PT substances to kill cancerous cells -
XX
PS Example 1; Page 15; 26pp; English.
XX
CC The present sequence is a peptide derived from human second
CC mitochondria derived activator of caspase (smac), also known as
CC direct inhibitor of apoptosis binding protein with low pI
CC (DIABLO), but with the native N-terminal alanine residue (see
CC ABB76209) acetylated. Claimed smac-derived peptides (see
CC ABB76208-19) bind to the Bir2 and Bir3 domain of XIAP, an
CC inhibitor of apoptosis protein (IAP) family member. Modification
CC of the N-terminal alanine destroys all binding affinity for the
CC protein. Thus, Kd values for Bir-3 and Bir-2 were each over 1,000
CC uM for the present peptide, compared with 0.43 +/- 0.06 uM and 6.0
CC +/- 0.9 uM, respectively, for the corresponding wild-type peptide.
CC The claimed smac-derived peptides can be used to identify candidate
CC substances which induce or promote apoptosis in cells. The assay
CC involves determination of the ability of candidate compounds to
CC disrupt the binding interaction between a smac peptide and an IAP
CC family member.
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 33; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 7
ABB76228
ID ABB76228 standard; Peptide; 10 AA.
XX
AC ABB76228;
XX
DT 09-AUG-2002 (first entry)
XX
DE Fluorosceinated smac (DIABLO) derived peptide.
XX
KW DIABLO; smac; inhibitor of apoptosis protein; IAP; apoptosis;
KW human; cancer; cytostatic; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal fluorescein"
XX
PN WO200230959-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US32121.
XX
PR 13-OCT-2000; 2000US-0687549.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Fesik SW, Meadows RP, Betz SP, Liu Z, Olejniczak ET, Sun C;
XX
DR WPI; 2002-444169/47.
XX
PT Novel peptide derived from wild-type human second mitochondria derived
PT activator of caspase protein useful for identifying candidate
PT substances to kill cancerous cells -
XX
PS Example 1; Page 14; 26pp; English.

```
XX
CC   The present sequence corresponds to amino acids 1-9 of human
CC   second mitochondria derived activator of caspase (smac), also known
CC   as direct inhibitor of apoptosis binding protein with low pI
CC   (DIABLO), but is fluorosceinated.  The peptide was used in a
CC   fluorescence polarisation-based competition assay designed to
CC   determine the binding affinity of variant smac peptides (see
CC   ABB76206-27) to the Bir-3 and Bir-2 domains of XIAP, an inhibitor
CC   of apoptosis protein (IAP) family member.  Claimed smac-derived
CC   peptides can be used to identify candidate substances which induce
CC   or promote apoptosis in cells.  The assay involves determination of
CC   the ability of candidate compounds to disrupt the binding
CC   interaction between a smac peptide and an IAP family member.
XX
SQ   Sequence   10 AA;

  Query Match             100.0%;  Score 33;  DB 23;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.26;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 8
ABP71314
ID   ABP71314 standard; peptide; 15 AA.
XX
AC   ABP71314;
XX
DT   28-APR-2003  (first entry)
XX
DE   Human Smac protein N-terminal fragment.
XX
KW   Omi; HtrA2; serine protease; inhibitor of apoptosis protein; IAP;
KW   caspase; apoptosis; cytostatic; immunosuppressive; neuroprotective;
KW   vasotropic; gene therapy; Smac.
XX
OS   Homo sapiens.
XX
PN   WO2003006680-A2.
XX
PD   23-JAN-2003.
XX
PF   15-JUL-2002; 2002WO-US22658.
XX
PR   13-JUL-2001; 2001US-305378P.
PR   14-DEC-2001; 2001US-340163P.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Alnemri ES;
XX
DR   WPI; 2003-221760/21.
XX
PT   New Omi nucleic acids and peptides that bind to an inhibitor of
PT   apoptosis proteins, useful for regulating or altering caspase-mediated
PT   apoptosis and for treating cancer, tumor, or autoimmune diseases -
XX
PS   Example 2; Fig 6; 83pp; English.
XX
CC   The invention relates to polynucleotides encoding an Omi (serine
CC   protease) peptide or polypeptide. The Omi peptide specifically binds to a
CC   portion of an Inhibitor of Apoptosis Protein (IAP). The Omi polypeptide
CC   induces caspase-independent apoptosis, or fails to have serine protease
CC   activity. The Omi peptides are useful for regulating or altering
CC   apoptosis, specifically caspase-mediated apoptosis, and as immunogens for
CC   raising antibodies. Enhancers of apoptosis are useful for treating
CC   cancers, tumours or for destroying cells that mediate autoimmune
CC   diseases. Compositions may also be used for the treatment of diseases
CC   associated with inappropriate activation of apoptosis such as
CC   neurodegenerative diseases and ischaemic injury. The antibodies can be
CC   used in isolating Omi peptides, polypeptides and their variants, in
CC   identifying molecules that interact with Omi peptides and polypeptides,
CC   and in inhibiting or enhancing the biological activity of Omi peptides
CC   and polypeptides. Sequences ABP71310-315 represent fragments of various
CC   IAP-binding proteins, used to determine Omi as a IAP-binding protein.
XX
SQ   Sequence   15 AA;

  Query Match             100.0%;  Score 33;  DB 24;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.42;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 9
ABB76208
ID   ABB76208 standard; Peptide; 20 AA.
XX
AC   ABB76208;
XX
DT   09-AUG-2002  (first entry)
XX
DE   Human smac (DIABLO) derived peptide.
XX
KW   DIABLO; smac; inhibitor of apoptosis protein; IAP; apoptosis;
KW   human; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   20
FT                   /note= "optional C-terminal protecting group"
XX
PN   WO200230959-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US32121.
XX
PR   13-OCT-2000; 2000US-0687549.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Fesik SW,  Meadows RP,  Betz SP,  Liu Z,  Olejniczak ET,  Sun C;
XX
DR   WPI; 2002-444169/47.
XX
PT   Novel peptide derived from wild-type human second mitochondria derived
PT   activator of caspase protein useful for identifying candidate
PT   substances to kill cancerous cells -
XX
PS   Claim 5; Page 7; 26pp; English.
XX
CC   The present sequence is a peptide derived from wild-type human
CC   second mitochondria derived activator of caspase (smac), also known
CC   as direct inhibitor of apoptosis binding protein with low pI
CC   (DIABLO).  The peptide is one of 12 claimed smac (DIABLO)-derived
CC   peptides (see ABB76208-19) which bind to the Bir2 and Bir3 domain
CC   of XIAP, an inhibitor of apoptosis protein (IAP) family member.
CC   Kd values for Bir-3 and Bir-2 are 0.69 +/- 0.05 uM and 6.7 +/- 0.7
CC   uM, respectively, for the present peptide, compared with 0.42 +/-
CC   0.02 uM and 2.3 +/- 0.3 uM, respectively, for full-length smac.
CC   Modification of the N-terminal alanine destroys binding affinity to
CC   XIAP.  For example, N-terminal acetylation of the present peptide,
CC   replacement of the N-terminal alanine with glycine, propionic acid
CC   or isobityric acid all resulted in Kd values for Bir-3 and for Bir-2
CC   of over 1,000 uM.  The claimed peptides can be used to identify
CC   candidate substances which induce or promote apoptosis in cells.
CC   The assay involves determination of the ability of candidate
CC   compounds to disrupt the binding interaction between a smac (DIABLO)
```

```
CC   peptide and an IAP family member.
XX
SQ   Sequence   20 AA;

  Query Match             100.0%;  Score 33;  DB 23;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 0.57;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 10
AAU78435
ID   AAU78435 standard; Peptide; 30 AA.
XX
AC   AAU78435;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Inhibitor of apoptosis (IAP) protein Smac, mutant Smac-N30.
XX
KW   Human; inhibitor of apoptosis; IAP; Smac; apoptosis; BID; BIR1; BIR2;
KW   Bcl2 interacting domain; caspase; BIR domain; BIR3; gene therapy;
KW   neoplastic cell; mutant; tumour.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200216418-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26492.
XX
PR   24-AUG-2000; 2000US-227735P.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Alnemri ES;
XX
DR   WPI; 2002-304115/34.
XX
PT   Novel Smac peptides and polynucleotides encoding the peptides, useful
PT   for stimulating apoptosis in neoplastic or tumour cell which
PT   overexpresses inhibitor of caspase, and for identifying apoptosis
PT   modulating compounds  -
XX
PS   Example 3; Fig 7; 78pp; English.
XX
CC   The invention relates to an isolated Smac peptide or polypeptide (I)
CC   and an isolated nucleic acid (II) encoding (I). Also described is a
CC   method of identifying a compound that inhibits apoptosis, comprising:
CC   (a) separately contacting several cell populations expressing a
CC   cytosolic Smac (a Smac isoform that begins with MKSDFYF sequence,
CC   replacing the mitochondrial targeting sequence (residues 1-55 of (I)),
CC   and residues 56-60 of (I)) and an inhibitor of BID (Bcl2 interacting
CC   domain) with a compound to be tested for apoptotic inhibiting activity;
CC   (b) incubating the cell populations with a direct stimulus of the cell
CC   death pathway; and (c) measuring the specific apoptotic activity of the
CC   cell populations, where inhibition of the specific apoptotic activity is
CC   indicative that the compound is an inhibitor of apoptosis. (I) and (II)
CC   are useful for inducing apoptosis in a cell. The Smac polypeptide and
CC   polynucleotide are useful for stimulating apoptosis in a neoplastic or
CC   tumour cell which overexpresses an inhibitor of caspase, where the
CC   inhibitor inhibits activation or activity of caspase-3, caspase-7 or
CC   caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
CC   (I) is useful for identifying an inhibitor or enhancer of a caspase-
CC   mediated apoptosis which involves contacting a cell transformed or
CC   transfected with a vector expressing (I) with a candidate inhibitor or
CC   candidate enhancer; and detecting cell viability, where an increase in
CC   cell viability indicates the presence of an inhibitor and a decrease in
CC   cell viability indicates the presence of an enhancer. Optionally, the
CC   method involves detecting the presence of large and small caspase
CC   subunits after contacting cell transformed with the vector expressing
CC   (I), with the candidate compound. A decrease in processing indicates the
CC   presence of an inhibitor and an increase in the processing indicates the
CC   presence of an enhancer. Preferably, the large and small subunits of
CC   caspase-3, caspase-7 or caspase-9 are detected. (I) is also useful for
CC   identifying a compound that inhibits Smac binding to Smac-binding
CC   molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
CC   or a full-length IAP). (II) is useful in gene therapy techniques. The
CC   present sequence represents the amino acid sequence of Smac mutant
CC   Smac-N30.
XX
SQ   Sequence   30 AA;

  Query Match             100.0%;  Score 33;  DB 23;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 0.91;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AVPIAQK 7
              |||||||
Db          1 AVPIAQK 7


RESULT 11
AAU78439
ID   AAU78439 standard; Peptide; 35 AA.
XX
AC   AAU78439;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Inhibitor of apoptosis (IAP) protein Smac, peptide Smac-N35.
XX
KW   Human; inhibitor of apoptosis; IAP; Smac; apoptosis; BID; BIR1; BIR2;
KW   Bcl2 interacting domain; caspase; BIR domain; BIR3; gene therapy;
KW   neoplastic cell; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200216418-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26492.
XX
PR   24-AUG-2000; 2000US-227735P.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Alnemri ES;
XX
DR   WPI; 2002-304115/34.
XX
PT   Novel Smac peptides and polynucleotides encoding the peptides, useful
PT   for stimulating apoptosis in neoplastic or tumour cell which
PT   overexpresses inhibitor of caspase, and for identifying apoptosis
PT   modulating compounds  -
XX
PS   Example 4; Page 47; 78pp; English.
XX
CC   The invention relates to an isolated Smac peptide or polypeptide (I)
CC   and an isolated nucleic acid (II) encoding (I). Also described is a
CC   method of identifying a compound that inhibits apoptosis, comprising:
CC   (a) separately contacting several cell populations expressing a
CC   cytosolic Smac (a Smac isoform that begins with MKSDFYF sequence,
CC   replacing the mitochondrial targeting sequence (residues 1-55 of (I)),
CC   and residues 56-60 of (I)) and an inhibitor of BID (Bcl2 interacting
CC   domain) with a compound to be tested for apoptotic inhibiting activity;
CC   (b) incubating the cell populations with a direct stimulus of the cell
CC   death pathway; and (c) measuring the specific apoptotic activity of the
CC   cell populations, where inhibition of the specific apoptotic activity is
CC   indicative that the compound is an inhibitor of apoptosis. (I) and (II)
```

CC are useful for inducing apoptosis in a cell. The Smac polypeptide and
CC polynucleotide are useful for stimulating apoptosis in a neoplastic or
CC tumour cell which overexpresses an inhibitor of caspase, where the
CC inhibitor inhibits activation or activity of caspase-3, caspase-7 or
CC caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
CC (I) is useful for identifying an inhibitor or enhancer of a caspase-
CC mediated apoptosis which involves contacting a cell transformed or
CC transfected with a vector expressing (I) with a candidate inhibitor or
CC candidate enhancer; and detecting cell viability, where an increase in
CC cell viability indicates the presence of an inhibitor and a decrease in
CC cell viability indicates the presence of an enhancer. Optionally, the
CC method involves detecting the presence of large and small caspase
CC subunits after contacting cell transformed with the vector expressing
CC (I), with the candidate compound. A decrease in processing indicates the
CC presence of an inhibitor and an increase in the processing indicates the
CC presence of an enhancer. Preferably, the large and small subunits of
CC caspase-3, caspase-7 or caspase-9 are detected. (I) is also useful for
CC identifying a compound that inhibits Smac binding to Smac-binding
CC molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
CC or a full-length IAP). (II) is useful in gene therapy techniques. The
CC present sequence represents the amino acid sequence of Smac peptide
CC Smac-N35.
XX
SQ Sequence 35 AA;

Query Match 100.0%; Score 33; DB 23; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 12
AAU78436
ID AAU78436 standard; Peptide; 39 AA.
XX
AC AAU78436;
XX
DT 18-JUN-2002 (first entry)
XX
DE Inhibitor of apoptosis (IAP) protein Smac, mutant Smac-N39.
XX
KW Human; inhibitor of apoptosis; IAP; Smac; apoptosis; BID; BIR1; BIR2;
KW Bcl2 interacting domain; caspase; BIR domain; BIR3; gene therapy;
KW neoplastic cell; mutant; tumour.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200216418-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26492.
XX
PR 24-AUG-2000; 2000US-227735P.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Alnemri ES;
XX
DR WPI; 2002-304115/34.
XX
PT Novel Smac peptides and polynucleotides encoding the peptides, useful
PT for stimulating apoptosis in neoplastic or tumour cell which
PT overexpresses inhibitor of caspase, and for identifying apoptosis
PT modulating compounds -
XX
PS Example 3; Fig 7; 78pp; English.
XX
CC The invention relates to an isolated Smac peptide or polypeptide (I)
CC and an isolated nucleic acid (II) encoding (I). Also described is a
CC method of identifying a compound that inhibits apoptosis, comprising:
CC (a) separately contacting several cell populations expressing a
CC cytosolic Smac (a Smac isoform that begins with MKSDFYF sequence,
CC replacing the mitochondrial targeting sequence (residues 1-55 of (I)),
CC and residues 56-60 of (I)) and an inhibitor of BID (Bcl2 interacting
CC domain) with a compound to be tested for apoptotic inhibiting activity;
CC (b) incubating the cell populations with a direct stimulus of the cell
CC death pathway; and (c) measuring the specific apoptotic activity of the
CC cell populations, where inhibition of the specific apoptotic activity is
CC indicative that the compound is an inhibitor of apoptosis. (I) and (II)
CC are useful for inducing apoptosis in a cell. The Smac polypeptide and
CC polynucleotide are useful for stimulating apoptosis in a neoplastic or
CC tumour cell which overexpresses an inhibitor of caspase, where the
CC inhibitor inhibits activation or activity of caspase-3, caspase-7 or
CC caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
CC (I) is useful for identifying an inhibitor or enhancer of a caspase-
CC mediated apoptosis which involves contacting a cell transformed or
CC transfected with a vector expressing (I) with a candidate inhibitor or
CC candidate enhancer; and detecting cell viability, where an increase in
CC cell viability indicates the presence of an inhibitor and a decrease in
CC cell viability indicates the presence of an enhancer. Optionally, the
CC method involves detecting the presence of large and small caspase
CC subunits after contacting cell transformed with the vector expressing
CC (I), with the candidate compound. A decrease in processing indicates the
CC presence of an inhibitor and an increase in the processing indicates the
CC presence of an enhancer. Preferably, the large and small subunits of
CC caspase-3, caspase-7 or caspase-9 are detected. (I) is also useful for
CC identifying a compound that inhibits Smac binding to Smac-binding
CC molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
CC or a full-length IAP). (II) is useful in gene therapy techniques. The
CC present sequence represents the amino acid sequence of Smac mutant
CC Smac-N39.
XX
SQ Sequence 39 AA;

Query Match 100.0%; Score 33; DB 23; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 13
AAU78430
ID AAU78430 standard; Peptide; 40 AA.
XX
AC AAU78430;
XX
DT 18-JUN-2002 (first entry)
XX
DE Inhibitor of apoptosis (IAP) protein Smac, N-terminal peptide.
XX
KW Human; inhibitor of apoptosis; IAP; Smac; apoptosis; BID; BIR1; BIR2;
KW Bcl2 interacting domain; caspase; BIR domain; BIR3; gene therapy;
KW neoplastic cell; tumour.
XX
OS Homo sapiens.
XX
PN WO200216418-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26492.
XX
PR 24-AUG-2000; 2000US-227735P.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI Alnemri ES;
XX

DR WPI; 2002-304115/34.
XX
PT Novel Smac peptides and polynucleotides encoding the peptides, useful
PT for stimulating apoptosis in neoplastic or tumour cell which
PT overexpresses inhibitor of caspase, and for identifying apoptosis
PT modulating compounds -
XX
PS Example 3; Fig 7; 78pp; English.
XX
CC The invention relates to an isolated Smac peptide or polypeptide (I)
CC and an isolated nucleic acid (II) encoding (I). Also described is a
CC method of identifying a compound that inhibits apoptosis, comprising:
CC (a) separately contacting several cell populations expressing a
CC cytosolic Smac (a Smac isoform that begins with MKSDFYF sequence,
CC replacing the mitochondrial targeting sequence (residues 1-55 of (I)),
CC and residues 56-60 of (I)) and an inhibitor of BID (Bcl2 interacting
CC domain) with a compound to be tested for apoptotic inhibiting activity;
CC (b) incubating the cell populations with a direct stimulus of the cell
CC death pathway; and (c) measuring the specific apoptotic activity of the
CC cell populations, where inhibition of the specific apoptotic activity is
CC indicative that the compound is an inhibitor of apoptosis. (I) and (II)
CC are useful for inducing apoptosis in a cell. The Smac polypeptide and
CC polynucleotide are useful for stimulating apoptosis in a neoplastic or
CC tumour cell which overexpresses an inhibitor of caspase, where the
CC inhibitor inhibits activation or activity of caspase-3, caspase-7 or
CC caspase-9. Preferably, the cell overexpresses at least a portion of IAP.
CC (I) is useful for identifying an inhibitor or enhancer of a caspase-
CC mediated apoptosis which involves contacting a cell transformed or
CC transfected with a vector expressing (I) with a candidate inhibitor or
CC candidate enhancer; and detecting cell viability, where an increase in
CC cell viability indicates the presence of an inhibitor and a decrease in
CC cell viability indicates the presence of an enhancer. Optionally, the
CC method involves detecting the presence of large and small caspase
CC subunits after contacting cell transformed with the vector expressing
CC (I), with the candidate compound. A decrease in processing indicates the
CC presence of an inhibitor and an increase in the processing indicates the
CC presence of an enhancer. Preferably, the large and small subunits of
CC caspase-3, caspase-7 or caspase-9 are detected. (I) is also useful for
CC identifying a compound that inhibits Smac binding to Smac-binding
CC molecule (a portion of IAP e.g. a BIR domain such as BIR1, BIR2 or BIR3,
CC or a full-length IAP). (II) is useful in gene therapy techniques. The
CC present sequence represents the N-terminal amino acid sequence of Smac
CC protein.
XX
SQ Sequence 40 AA;

Query Match 100.0%; Score 33; DB 23; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 1 AVPIAQK 7

RESULT 14
ABG72303
ID ABG72303 standard; Protein; 84 AA.
XX
AC ABG72303;
XX
DT 29-JAN-2003 (first entry)
XX
DE Rat partial sequence for pro-apoptotic protein DIABLO.
XX
KW Rat; pro-apoptotic protein; DIABLO; cell death; apoptosis;
KW inhibitor of apoptosis; IAP; cancer; vascular disease; hepatic disease;
KW autoimmune disease; neurodegenerative disease; tissue damage;
KW muscular tissue damage; heart attack; hepatic tissue damage;
KW liver disease; immunogen.
XX
OS Rattus sp.
XX
PN US2002110851-A1.
XX
PD 15-AUG-2002.
XX
PF 02-MAR-2001; 2001US-0798116.
XX
PR 02-MAR-2000; 2000AU-0005995.
XX
PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI Verhagen AM, Ekert PG, Vaux DL;
XX
DR WPI; 2003-074681/07.
XX
PT New pro-apoptotic polypeptide, useful for screening for agents which
PT modulate cell death and for treating conditions associated with cell
PT death or apoptosis e.g. cancer -
XX
PS Disclosure; Page 35; 50pp; English.
XX
CC The invention relates to an isolated pro-apoptotic polypeptide,
CC designated DIABLO, or its biologically active fragment of 8 amino acids
CC in length. Also included are the polynucleotide encoding DIABLO,
CC expression vectors, transformed host cells, producing a biologically
CC active fragment of DIABLO (by contacting an inhibitor of apoptosis (IAP)
CC with a fragment of the polypeptide, and detecting a reduction in activity
CC of the IAP), producing a natural or synthetic variant of DIABLO
CC with cell death activity or which reduces IAP activity, an antigen-
CC binding molecule that specifically binds to DIABLO or its fragment,
CC detecting DIABLO in a biological sample (by contacting the sample
CC with an IAP and detecting the presence of an IAP/DIABLO complex),
CC modulating the death of a cell (by contacting a cell with an
CC agent, which modulates the level and/or functional activity of a
CC polypeptide), a composition for treatment/prophylaxis of a DIABLO related
CC condition comprising an agent which reduces the level/activity of a
CC polypeptide or DIABLO. DIABLO, or a nucleic acid encoding DIABLO, is
CC useful for screening for an agent which modulates cell death. An
CC antigen-binding molecule is useful for detecting DIABLO in a biological
CC sample. The agent which modulates the level and/or functional activity of
CC a polypeptide comprising mature or pro-human DIABLO polypeptide, is
CC useful for the treatment and/or prophylaxis of a condition associated
CC with expression or activation of DIABLO, such as cancer, vascular
CC disease, hepatic disease, autoimmune disease and neurodegenerative
CC disease, tissue damage or muscular tissue damage associated with heart
CC attack, or hepatic tissue damage associated with a liver disease.
CC DIABLO is also useful for treatment and/or prophylaxis of conditions
CC associated with cell death or apoptosis. The present sequence
CC represents partial rat DIABLO.
XX
SQ Sequence 84 AA;

Query Match 100.0%; Score 33; DB 24; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 54 AVPIAQK 60

RESULT 15
ABG72302
ID ABG72302 standard; Protein; 202 AA.
XX
AC ABG72302;
XX
DT 29-JAN-2003 (first entry)
XX
DE Human partial sequence for pro-apoptotic protein DIABLO.
XX
KW Human; pro-apoptotic protein; DIABLO; cell death; apoptosis;
KW inhibitor of apoptosis; IAP; cancer; vascular disease; hepatic disease;
KW autoimmune disease; neurodegenerative disease; tissue damage;

KW muscular tissue damage; heart attack; hepatic tissue damage;
KW liver disease; immunogen.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT sig_peptide 1..25
FT /partial
FT mat_peptide 26..202
FT /label= Mature_DIABLO
XX
PN US2002110851-A1.
XX
PD 15-AUG-2002.
XX
PF 02-MAR-2001; 2001US-0798116.
XX
PR 02-MAR-2000; 2000AU-0005995.
XX
PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI Verhagen AM, Ekert PG, Vaux DL;
XX
DR WPI; 2003-074681/07.
XX
PT New pro-apoptotic polypeptide, useful for screening for agents which
PT modulate cell death and for treating conditions associated with cell
PT death or apoptosis e.g. cancer -
XX
PS Disclosure; Fig 2E; 50pp; English.
XX
CC The invention relates to an isolated pro-apoptotic polypeptide,
CC designated DIABLO, or its biologically active fragment of 8 amino acids
CC in length. Also included are the polynucleotide encoding DIABLO,
CC expression vectors, transformed host cells, producing a biologically
CC active fragment of DIABLO (by contacting an inhibitor of apoptosis (IAP)
CC with a fragment of the polypeptide, and detecting a reduction in activity
CC of the IAP), producing a natural or synthetic variant of DIABLO
CC with cell death activity or which reduces IAP activity, an antigen-
CC binding molecule that specifically binds to DIABLO or its fragment,
CC detecting DIABLO in a biological sample (by contacting the sample
CC with an IAP and detecting the presence of an IAP/DIABLO complex),
CC modulating the death of a cell (by contacting a cell with an
CC agent, which modulates the level and/or functional activity of a
CC polypeptide), a composition for treatment/prophylaxis of a DIABLO related
CC condition comprising an agent which reduces the level/activity of a
CC polypeptide or DIABLO. DIABLO, or a nucleic acid encoding DIABLO, is
CC useful for screening for an agent which modulates cell death. An
CC antigen-binding molecule is useful for detecting DIABLO in a biological
CC sample. The agent which modulates the level and/or functional activity of
CC a polypeptide comprising mature or pro-human DIABLO polypeptide, is
CC useful for the treatment and/or prophylaxis of a condition associated
CC with expression or activation of DIABLO, such as cancer, vascular
CC disease, hepatic disease, autoimmune disease and neurodegenerative
CC disease, tissue damage or muscular tissue damage associated with heart
CC attack, or hepatic tissue damage associated with a liver disease.
CC DIABLO is also useful for treatment and/or prophylaxis of conditions
CC associated with cell death or apoptosis. The present sequence
CC represents partial human DIABLO.
XX
SQ Sequence 202 AA;

Query Match 100.0%; Score 33; DB 24; Length 202;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVPIAQK 7
|||||||
Db 19 AVPIAQK 25

Search completed: September 12, 2003, 11:13:24
Job time : 84 secs

(Rev)